

Máster Interuniversitario en Estadística e Investigación Operativa UPC-UB

Título: El Uso Del Propensity Score Matching En Estudios Observacionales

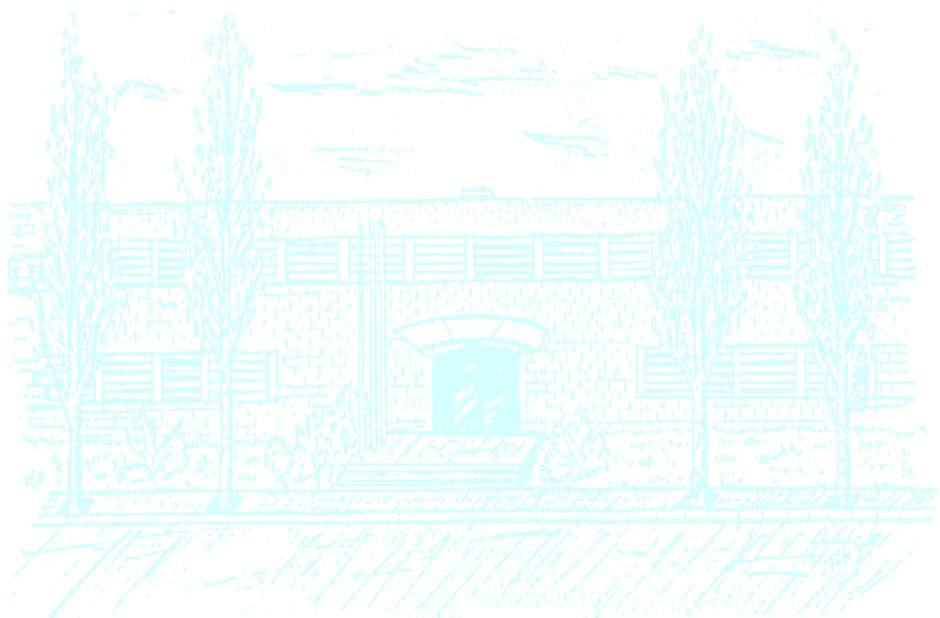
Autor: Nieto Cerna, Wagner Roger

Director: Langohr, Klaus Gerhard

Departamento: Estadística e Investigación Operativa

Universidad: Universidad Politécnica de Cataluña (UPC)

Convocatoria: Junio de 2020



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



AGRADECIMIENTOS

A Klaus Langohr, por tutorizarme este proyecto.

A la Universidad Politécnica de Cataluña, por su formación.

A mis profesores, les doy las más eternas gracias, por haberse preocupado, por regañarme y ser duros cuando no hice las cosas bien, por ir más allá de instruirnos.

A mis hermanos y a todo el resto de mi familia, hacia quienes sólo puedo expresar mi sincero agradecimiento por su apoyo y comprensión.

Y muy especialmente, a Fanny Hernández, por su amor y apoyo incondicional, por estar ahí desde el principio hasta el final, por su inagotable paciencia.

ÍNDICE

1.	INTRODUCCIÓN	5
1.1.	OBJETIVOS	5
1.2.	JUSTIFICACIÓN Y ALCANCE	6
2.	MARCO TEÓRICO DE LA INFERENCIA CAUSAL	7
2.1.	UNA DEFINICIÓN DE CAUSALIDAD	7
2.2.	DEFINICIÓN DEL EFECTO CAUSAL	8
2.2.1.	Efectos causales individuales	8
2.2.2.	Efectos causales promedio	10
2.3.	MEDIDAS DE EFECTO CAUSAL	12
2.4.	VARIABILIDAD ALEATORIA	12
2.5.	CAUSALIDAD VERSUS ASOCIACIÓN	14
2.6.	EXPERIMENTOS ALEATORIOS	16
2.6.1.	Aleatorización	18
2.6.2.	Aleatorización condicional	19
2.6.3.	Estandarización	20
2.6.4.	Probabilidad inversa de ponderación	22
2.7.	ESTUDIOS OBSERVACIONALES	27
2.7.1.	Condiciones de identificabilidad	27
2.7.2.	Intercambiabilidad	28
2.7.3.	Positividad	29
3.	PROPENSITY SCORE	30
3.1.	DEFINICIÓN DEL PROPENSITY SCORE	31
3.1.1.	Propiedades del propensity score	32
3.1.2.	Estimación del propensity score	32
3.1.3.	Selección del modelo del propensity score	33
3.1.4.	Modelo de regresión logística	35
3.1.5.	Estimación del propensity score mediante regresión logística	36
3.1.6.	Otros métodos para estimar el propensity score	37
3.1.6.1.	Modelos Probit	37
3.1.6.2.	Análisis discriminante	37
3.1.6.3.	Clasificador Bayesiano	38
3.1.6.4.	Bosques aleatorios	39
3.2.	MÉTODOS DEL PROPENSITY SCORE	39
3.2.1.	Matching por el propensity score	41
3.2.2.	Diferentes métodos de matching	49
3.2.2.1.	Estudio de la simulación de los diferentes métodos PSM	49
3.2.2.2.	Verificación de equilibrio en la distribución de covariables	49
3.2.2.3.	Verificación de equilibrio mediante gráficos	49
3.2.2.4.	Verificación de equilibrio con pruebas estadísticas	50

3.2.2.5.	Verificación de equilibrio con diferencias estandarizadas	50
3.2.2.6.	Estimación de efectos de tratamiento basado en matching.....	51
3.2.3.	Estratificación por el propensity score.....	52
3.2.4.	Ponderación por el propensity score.....	55
3.2.5.	Propensity score como variable de ajuste	57
4.	APLICACIÓN DEL PROPENSITY SCORE PARA LA EVALUACIÓN DEL RENDIMIENTO NEUROPSICOLÓGICO EN USUARIOS DE CANNABIS	59
4.1.	EL ESTUDIO DE VERDEJO ET AL. (2013)	59
4.2.2.	Objetivos del estudio	59
4.2.3.	Materiales y métodos	60
4.2.4.	Procedimiento de la prueba	61
4.2.5.	Evaluación neuropsicológica.....	61
4.2.6.	Análisis estadístico	62
4.2.	APLICACIÓN DEL PROPENSITY SCORE.....	63
4.2.1.	Missing data	63
4.2.2.	Estimación del propensity score.....	64
4.2.3.	Matching por el propensity score	65
4.2.4.	Verificación de equilibrio en la distribución de covariables	66
4.2.5.	Verificación de equilibrio mediante gráficos	66
4.2.6.	Verificación de equilibrio con pruebas estadísticas.....	71
4.2.7.	Verificación de equilibrio con diferencias estandarizadas.....	72
4.2.8.	Ajuste de la muestra emparejada en el modelo ANCOVA	73
4.2.9.	Estratificación por el propensity score.....	75
4.2.9.1.	Verificación de equilibrio mediante gráficos	76
4.2.9.2.	Verificación de equilibrio con pruebas estadísticas.....	78
4.2.9.3.	Verificación de equilibrio con diferencias estandarizadas.....	79
4.2.9.4.	Ajuste de la muestra estratificada en el modelo ANCOVA.....	80
4.2.10.	Ponderación por el propensity score	81
4.2.11.	Ajuste de la muestra ponderada en el modelo ANOVA	81
4.2.12.	El propensity score como variable de ajuste en el modelo ANCOVA	83
5.	DISCUSIÓN Y CONCLUSIONES	86
5.1.	DISCUSIÓN.....	86
5.2.	CONCLUSIONES.....	87
6.	BIBLIOGRAFÍA.....	88
7.	APÉNDICE	91
7.1.	APÉNDICE 1	91
7.2.	APÉNDICE 2	95
7.3.	APÉNDICE 3.....	104
7.4.	APÉNDICE 4.....	117

1. INTRODUCCIÓN

En los ensayos clínicos aleatorizados, la asignación del tratamiento genera dos o más muestras equilibradas, garantizando la ausencia de sesgo de confusión entre los individuos asignados a los grupos de tratamiento. Sin embargo, los ensayos clínicos aleatorizados no siempre son realizables, ya que, los individuos deben cumplir algunos requisitos para la realización de experimentos. Principalmente deben satisfacer los puntos de partida de la ética para la realización de “experimentos médico permisibles” internacionalmente reconocida como los principios del código de Núremberg (Clotet y Videla, 2003). En situaciones no factibles para la realización de estudios experimentales, solo podemos soñar con el diseño de estudios observacionales utilizando datos de observación.

En 1945, Cochran, definió un estudio observacional como una investigación empírica en la que el objetivo es dilucidar las relaciones causales en entornos en los que no es factible experimentación controlada. Identificar el verdadero efecto del tratamiento a partir de los datos de observación es un desafío porque solo podemos observar el resultado correspondiente al tratamiento recibido por un individuo, mientras que los resultados bajo tratamientos alternativos no se observan, esto se llama el problema fundamental de la inferencia causal. La limitación más importante de los estudios observacionales está en que los mecanismos de asignación del tratamiento que generalmente dependen de las características del individuo, no son conocidos ni aleatorios. Por lo tanto, existe un sesgo de selección por indicación de tratamiento, además las muestras generadas no se equilibran. En la literatura existen diferentes métodos propuestos para generar muestras equilibradas y minimizar el sesgo de selección existente en los estudios observacionales, uno de los métodos más poderosos y potentes es el uso del propensity score (Rosenbaum, P.R., 1983).

1.1. OBJETIVOS

Los objetivos del presente trabajo final de máster son los siguientes.

Principales:

1. Describir y evaluar cuatro métodos vanguardistas del propensity score en estudios observacionales, para reducir el efecto de las variables confusoras.
2. Una comprensión completa del propensity score. Por qué se usa, cómo se usa y para qué se usa.
3. Comparar los resultados obtenidos de un estudio real con los diferentes métodos del propensity score.

Secundarios:

1. Identificar las limitaciones y fortalezas de las diferentes variantes del propensity score.
2. Analizar la calidad de los diferentes métodos de *matching* y estratificación por el propensity score.

1.2. JUSTIFICACIÓN Y ALCANCE

El problema crucial en la estimación del efecto causal de los datos de observación se denomina sesgo de confusión, por ejemplo, es probable que los pacientes que reciben diferentes tratamientos entre sí exhiban diferentes características de pretratamiento que terminen afectando el desenlace final en los resultados. Normalmente, para la estimación del efecto de un tratamiento con datos de estudios observacionales usamos métodos estadísticos elegantes de análisis multivariable para controlar las diferencias existentes entre los individuos tratados y no tratados.

Los métodos clásicos para el control de las variables de confusión son restringidos al número de variables de ajuste, ya que el modelo debe seguir el principio de parsimonia, además, para variables de alta dimensionalidad realizar ajustes en modelos estadísticos resulta imposible, además, cuando son factibles de realizar el ajuste para variable de gran dimensión, los resultados obtenidos son difíciles de interpretar.

En este contexto, Rosenbaum y Rubin, en 1983, presentaron por primera vez un método estadístico revolucionario, una maravilla distintiva, conocida como el propensity score que sirve para condensar y resumir la información con variables de alta dimensionalidad y controlar los factores de confusión. El análisis del propensity score se desarrolló originalmente en datos trasversales, a medida que los fenómenos de investigación se han vuelto multifacéticos y multidimensionales, los datos de investigación correspondientes se han vuelto cada vez más complicados.

La complejidad de los datos de investigación plantea desafíos metodológicos para el desarrollo y uso del análisis del propensity score. Uno de los usos del propensity score es para generar una muestra “pseudoaleatorio”, a partir de los datos de observación.

La muestra estará formada por individuos tratados y no tratados, con el mismo valor en el propensity score, entonces bajo el supuesto de que la asignación del tratamiento en ambos grupos de tratamiento se ha realizado aleatoriamente, en el sentido de que cada individuo es igualmente propenso a recibir el tratamiento, los individuos tratados y no tratados tendrán una distribución similar en las covariables observadas.

El presente proyecto está estructurado en siete capítulos: En el Capítulo 2 se presenta el marco teórico de la inferencia causal, es decir, conceptos y métodos para la identificación no paramétrica de los efectos causales, el Capítulo 3 trata sobre el propensity score (es decir, consiste en un estudio bibliográfico exhaustiva sobre las bases teóricas y los diferentes métodos del propensity score), el Capítulo 4 trata sobre la aplicación del propensity score en datos de estudios de cohorte y caso-control con los resultados de su implementación en R, el Capítulo 5 trata de la discusión y conclusiones de los resultados finales del proyecto, en el Capítulo 6 se presenta las bibliografías utilizadas a lo largo del desarrollo de este proyecto y en el Capítulo 7 se presenta los códigos de la implementación en R.

2. MARCO TEÓRICO DE LA INFERENCIA CAUSAL

Este capítulo es una breve introducción de los conceptos fundamentales de la inferencia causal y las terminologías de causalidad, asociación y efectos causales. En general, las definiciones, ejemplos y notaciones matemáticas que usamos, son de la literatura introducidas por Hernán y Robins (2020) y las transparencias de la asignatura de Klaus Langohr (Curso 2018/19: Epidemiología).

La importancia de estudiar relaciones causales, la red causal que vincula el tratamiento, el resultado y otras variables es para ayudar a tomar mejores decisiones a los científicos, especialmente a los científicos sociales y de la salud.

En el campo de las ciencias empíricas como la física, las relaciones causales ayudan generar conocimiento y aportar explicaciones. Por ejemplo, cuando un físico cuántico computa la probabilidad de que un átomo absorba un protón, está determinando la causa por la que el átomo salta a un nivel de energía, en otras palabras, intenta establecer una relación de causa y efecto

2.1. UNA DEFINICIÓN DE CAUSALIDAD

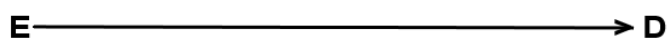
Definición de causalidad En epidemiología, la causalidad se puede definir de la siguiente manera (Doménech y Granero 2000): Una exposición E puede considerarse un factor de riesgo de un resultado D , si se cumplen las siguientes tres condiciones:

1. La exposición debe preceder al resultado.
2. Debe haber una asociación significativa entre E y D .
3. La asociación entre E y D no debe ser el resultado de una asociación entre una tercera variable y ambos D y E .

Enfermedad y exposición: estructuras causales

La pregunta decisiva es si una asociación estadísticamente significativa entre E y D es, en realidad, debido a la relación causal entre ambos o no. Según Hernán et al (2004), una asociación entre E y D puede ser el resultado de tres estructuras causales diferentes:

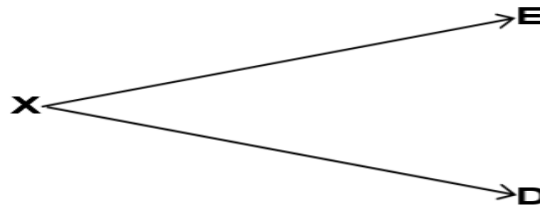
1. **Causa y efecto** La exposición E , en realidad, causa el resultado D .



Por ejemplo, se ha demostrado que fumar es un factor de riesgo para varios tipos de cáncer, o que el tratamiento antirretroviral previene el SIDA.

2. Causas comunes de exposición y resultado.

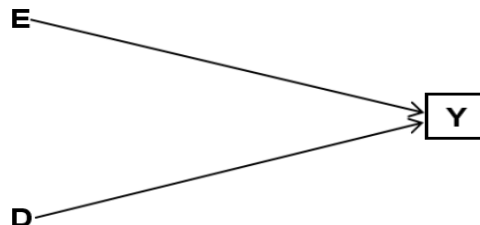
Si E y D comparten una causa común, normalmente están asociadas.



Por ejemplo, la asociación positiva entre altura e inteligencia entre los niños, X es la edad.

3. Efectos comunes de la exposición y el resultado

Si E y D tienen un efecto común Y , ambos están asociados condicionalmente dentro de los niveles de Y .



Por ejemplo, E podría ser un gen de interés y D un trastorno mental. Si ambos tienen un efecto común, como el consumo de drogas, uno puede detectar una asociación entre E y D , si se recopilan datos entre las personas que consumen drogas.

2.2. DEFINICIÓN DEL EFECTO CAUSAL

A continuación, vamos a introducir la definición del efecto causal, primero en un individuo y luego en la población de individuos.

2.2.1. Efectos causales individuales

Para definir el efecto causal necesitamos introducir algunas notaciones para conceptualizar en símbolos matemáticos.

Algunas notaciones

A continuación, usamos la notación introducida en el ejemplo del libro de Hernán y Robins (2020).

- Variable de tratamiento dicotómico: A (1: tratado; 0: no tratado)
- Variable de resultado dicotómico: Y (1: muerte; 0: supervivencia)
- $Y^{a=i}$: Resultado bajo tratamiento $a = i$, $i \in \{0, 1\}$.

Definición del efecto causal para un individuo: En un individuo, el tratamiento A tiene un efecto causal en el resultado Y si

$$Y^{a=1} \neq Y^{a=0}$$

Si los resultados bajo el tratamiento y no tratamiento son diferentes, entonces existe un efecto causal en el resultado.

EJEMPLO 1

Zeus es un paciente que espera un trasplante de corazón. El 1 de enero, recibe un nuevo corazón. Cinco días después, él muere. Y supongamos que de alguna manera sabemos que, si Zeus no hubiera recibido un trasplante de corazón el 1 de enero, habría estado vivo cinco días después. Entonces el trasplante de corazón causó la muerte de Zeus.

Zeus: $Y^{a=1} = 1 \neq 0 = Y^{a=0} \Rightarrow$ el tratamiento tiene un efecto causal

Por lo tanto, la intervención de trasplante de corazón tuvo un efecto causal en la supervivencia de cinco días de Zeus.

EJEMPLO 2

Otra paciente, Hera, también recibió un trasplante de corazón el 1 de enero. Cinco días después estaba viva. Imaginemos que de alguna manera podemos saber que, si Hera no hubiera recibido el corazón el 1 de enero, todavía habría estado viva cinco días después. Entonces el trasplante de corazón no causó la supervivencia de Hera.

Hera: $Y^{a=1} = Y^{a=0} = 0 \Rightarrow$ el tratamiento no tiene efecto causal

Por lo tanto, el trasplante no tuvo un efecto causal en la supervivencia de cinco días de Hera.

Importante:

- Solo uno de los dos resultados bajo tratamiento es observable, el otro es su contrafactual o resultado potencial.
- En la vida real, los efectos causales individuales solo se pueden estimar bajo ciertas condiciones.

2.2.2. Efectos causales promedio

Para conceptualizar los efectos causales promedio, necesitamos tres piezas de información: un resultado de interés, las acciones $a = 1$ y $a = 0$ para comparar, y una población bien definida de individuos cuyos resultados $Y^{a=0}$ y $Y^{a=1}$ son comparables.

Definición del efecto causal promedio en la población: Un efecto causal promedio del tratamiento A sobre el resultado Y está presente en una población de interés si los riesgos contrafactuales $Pr[Y^{a=1} = 1]$ y $Pr[Y^{a=0} = 1]$ son diferentes, en símbolos

$$Pr[Y^{a=1} = 1] \neq Pr[Y^{a=0} = 1]$$

Esta definición, quiere decir, que el tratamiento A tiene un efecto causal promedio en el resultado Y en nuestra población.

Los riesgos contrafactuales $Pr[Y^{a=1} = 1]$ y $Pr[Y^{a=0} = 1]$ para el ejemplo, caso Zeus:

- $Pr[Y^{a=1} = 1]$: Riesgo de muerte bajo tratamiento
- $Pr[Y^{a=0} = 1]$: Riesgo de muerte sin tratamiento

EJEMPLO

Consideremos la familia extendida de Zeus como nuestra población de interés. La Tabla 1 muestra los resultados contrafactuales tanto con tratamiento ($a = 1$) como sin tratamiento ($a = 0$) para los 20 miembros de nuestra población.

Tabla 1 Resultados contrafactuales de la familia extendida de Zeus (Tabla 1.1 en Hernán & Robins).					
	$Y^{a=0}$	$Y^{a=1}$		$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1	Leto	0	1
Kronos	1	0	Ares	1	1
Demeter	0	0	Athena	1	1
Hades	0	0	Hephaestus	0	1
Hestia	0	0	Aphrodite	0	1
Poseidon	1	0	Cyclope	0	1
Hera	0	0	Persephone	1	1
Zeus	0	1	Hermes	1	0
Artemis	1	1	Hebe	1	0
Apollo	1	0	Dionysus	1	0

Primero, centremos nuestra atención en la última columna: el resultado $Y^{a=1}$ que se habría observado para cada individuo si hubiera recibido el tratamiento (un trasplante de corazón).

La mitad de los miembros de la población (**10 de 20**) habrían muerto si hubieran recibido un trasplante de corazón. Es decir, la proporción de individuos que habrían desarrollado el resultado si todos los individuos de la población recibieran $a = 1$ es:

$$Pr[Y^{a=1} = 1] = \frac{10}{20} = 0.5.$$

De manera similar, de la otra columna de la Tabla 1, podemos concluir que la mitad de los miembros de la población (**10 de 20**) habrían muerto si no hubieran recibido un trasplante de corazón. Es decir, la proporción de individuos que habrían desarrollado el resultado si todos los individuos de la población recibieran $a = 0$ es:

$$Pr[Y^{a=0} = 1] = \frac{10}{20} = 0.5.$$

Bajo la definición establecida previamente, el tratamiento **A** no tiene un efecto causal promedio en el resultado **Y** en nuestra población porque tanto el riesgo de muerte bajo tratamiento $Pr[Y^{a=1} = 1]$ como el riesgo de muerte sin tratamiento $Pr[Y^{a=0} = 1]$ son **0.5**. Es decir, no importa si todos o ninguno de los individuos reciben un trasplante de corazón: la mitad de ellos moriría, en cualquier caso.

Por lo tanto, no hay un efecto causal promedio en la familia de Zeus ya que

$$Pr[Y^{a=1} = 1] = Pr[Y^{a=0} = 1] = \frac{10}{20} = 0.5$$

Eso no implica la ausencia de efectos individuales, ya que, la Tabla 1 muestra que el tratamiento tiene un efecto causal individual en **12** miembros (incluido Zeus) de la población porque, para cada uno de estos **12** individuos, el valor de sus resultados contrafactuales $Y^{a=1}$ y $Y^{a=0}$ son diferentes.

Para que la definición se aplique tanto a los resultados dicotómicos como a los no dicotómicos. Podemos reescribir la definición de un efecto causal promedio no nulo en la población como

$$E[Y^{a=1}] \neq E[Y^{a=0}]$$

Por ejemplo, a pesar de que no podamos estimar los efectos causales individuales, sí es posible, bajo ciertas premisas, estimar los efectos causales medios.

2.3. MEDIDAS DE EFECTO CAUSAL

El efecto causal medio se puede cuantificar de distintas maneras, por ejemplo, mediante la diferencia de riesgo causal, la razón de riesgo causal o el odds ratio causal. A continuación, se presenta formalmente las tres medidas del efecto causal medio:

- $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] = 0$ (Diferencia de riesgo causal)
- $\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]} = 1$ (Razón de riesgo causal)
- $\frac{\Pr[Y^{a=1}=1]/\Pr[Y^{a=1}=0]}{\Pr[Y^{a=0}=1]/\Pr[Y^{a=0}=0]} = 1$ (Odds ratio causal)

2.4. VARIABILIDAD ALEATORIA

Hasta ahora nuestro objetivo ha sido calcular el efecto causal individual y promedio del tratamiento A sobre el resultado Y , al realizar los cálculos, nuestras inferencias causales podrían estar equivocadas. Hay dos razones cualitativamente diferentes por las cuales las inferencias causales pueden estar equivocadas: sesgo sistemático y variabilidad aleatoria. A continuación, se analiza el caso de la variabilidad aleatoria y cómo la tratamos. Para una muestra extraída aleatoriamente de una población de interés, en inferencia causal, la variabilidad aleatoria (error aleatorio) tiene dos fuentes: la variabilidad del muestreo, los contrafactuales no deterministas o ambos.

Variabilidad de muestreo

Decimos que la variabilidad de muestreo es una fuente de error aleatorio cuando las probabilidades de la población de interés $\Pr[Y^{a=1} = 1]$ y $\Pr[Y^{a=0} = 1]$ no se pueden calcular, solo estimar por las proporciones de la muestra $\widehat{Pr}[Y^{a=1} = 1]$ y $\widehat{Pr}[Y^{a=0} = 1]$ respectivamente.

Por ejemplo, consideremos los individuos de la Tabla 1, visto como una población de veinte personas. Supongamos que los vemos como una muestra aleatoria de una superpoblación mucho más grande, casi infinita (por ejemplo, todos los inmortales). Denotamos la proporción de individuos en la muestra que habrían muerto si no hubieran sido tratados como $\widehat{Pr}[Y^{a=0} = 1] = 10/20 = 0.5$.

La proporción de los individuos en la muestra $\widehat{Pr}[Y^{a=0} = 1]$ no tiene que ser exactamente igual a la proporción de individuos que habrían muerto si la superpoblación completa no hubiera sido tratada, $\Pr[Y^{a=0} = 1]$. Por ejemplo, supongamos que $\Pr[Y^{a=0} = 1] = 0.57$ en la población, pero, debido a un error aleatorio en la variabilidad del muestreo, $\widehat{Pr}[Y^{a=0} = 1] = 0.5$ en nuestra muestra particular. Por lo tanto, disponiendo solo de: $\widehat{Pr}[Y^{a=1} = 1]$ y $\widehat{Pr}[Y^{a=0} = 1]$, no se puede afirmar con certeza que existe o no un efecto causal promedio. Por el contrario, se debe utilizar un procedimiento estadístico para evaluar la evidencia empírica con respecto a la hipótesis nula causal $\Pr[Y^{a=1} = 1] = \Pr[Y^{a=0} = 1]$.

Contrafactual no determinista

Decimos que la contrafactual no determinista es una fuente de error aleatorio cuando los resultados contrafactuales $Y^{a=1}$ y $Y^{a=0}$ de un individuo pueden ser no fijo, más bien estocástico.

Escenario 1:

Volvamos a nuestro ejemplo caso familia extendida de Zeus como nuestra población de interés, sobre el trasplante de corazón y mortalidad. Zeus habría muerto si hubiera sido tratado y habría sobrevivido si no hubiera sido tratado.

Según lo definido, los valores de los resultados contrafactuales son fijos o deterministas para cada individuo. Por ejemplo, para Zeus

Resultados contrafactuales fijos: $Y^{a=1} = 1$ y $Y^{a=0} = 0$.

En otra palabra, Zeus tiene un **100%** de posibilidades de morir si se trata y un **0%** de posibilidades de morir si no se trata.

Escenario 2

Ahora imaginemos otro escenario en el que Zeus tiene un **90%** de posibilidades de morir si se trata, y un **10%** de posibilidades de morir si no se trata. En este escenario, los resultados contrafactuales son estocásticos o no deterministas ya que las probabilidades de Zeus de morir bajo tratamiento (**0.9**) y sin tratamiento (**0.1**) no son cero ni uno.

Los valores de $Y^{a=1}$ y $Y^{a=0}$ que se muestran en la Tabla 1 serían posibles realizaciones de “lanzamientos aleatorios de monedas de mortalidad” con estas probabilidades. Además, estas probabilidades varían entre individuos puesto que no todos los individuos son igualmente susceptibles de desarrollar el resultado. Por lo tanto,

Resultados contra factuales estocásticos: $Y^{a=1}$ y $Y^{a=0}$.

También es importante destacar que la mecánica cuántica, en contraste con la mecánica clásica, sostiene que los resultados son inherentemente no deterministas. Es decir, si la probabilidad mecánica cuántica de la muerte de Zeus es del **90%**, la teoría sostiene que, sin importar la cantidad de datos que recopilemos sobre Zeus, la incertidumbre sobre si Zeus realmente desarrollará el resultado si se trata es irreducible.

2.5. CAUSALIDAD VERSUS ASOCIACIÓN

Los datos disponibles de los estudios reales se ven diferentes de los que se muestran en la Tabla 1, ya que, en el mundo real, solo podemos observar uno de los resultados contrafactuales. Nos referimos al resultado observado como Y . Así, para cada individuo, conocemos el nivel de tratamiento observado A y el resultado Y como en la Tabla 2.

Tabla 2: Resultados observables de la familia extendida de Zeus (Tabla 1.2 en Hernán & Robins).

	A	Y		A	Y
Rheia	0	0	Leto	0	0
Kronos	0	1	Ares	1	1
Demeter	0	0	Athena	1	1
Hades	0	0	Hephaestus	1	1
Hestia	1	0	Aphrodite	1	1
Poseidon	1	0	Cyclope	1	1
Hera	1	0	Persephone	1	1
Zeus	1	1	Hermes	1	0
Artemis	0	1	Hebe	1	0
Apollo	0	1	Dionysus	1	0

Los datos de la Tabla 2 se pueden usar para calcular la proporción de individuos que desarrollaron el resultado Y entre aquellos individuos de la población que recibieron el valor del tratamiento a .

Definición de asociación Decimos que el tratamiento A y el resultado Y son asociados o dependientes cuando

$$Pr[Y = 1|A = 1] \neq Pr[Y = 1|A = 0].$$

En otras palabras, la asociación se define por un riesgo diferente en dos subconjuntos disjuntos de la población determinados por el valor real del tratamiento de los individuos ($A = 1$ o $A = 0$).

Las inferencias sobre la asociación están relacionadas con preguntas en el mundo real, como "¿cuál es el riesgo en los tratados?" y "¿cuál es el riesgo en los no tratados?"

Para los resultados dicotómicos y continuos Y , la asociación en la población se define como

$$E[Y|A = 1] \neq E[Y|A = 0]$$

Donde, el riesgo es igual al promedio en la población, es decir,

$$Pr[Y = 1|A = 1] = E[Y|A = 1] , Pr[Y = 1|A = 0] = E[Y|A = 0]$$

EJEMPLO

En nuestra población, el tratamiento y el resultado están asociados porque,

$$Pr[Y = 1|A = 1] = 7/13 \text{ y } Pr[Y = 1|A = 0] = 3/7.$$

Por lo tanto, en nuestra población, hubo asociación porque el riesgo de mortalidad en los tratados (7/13) fue mayor que en los no tratados (3/7).

Medidas de asociación

Cuando existe asociación: la diferencia de riesgo asociativo, la razón de riesgo y la razón de probabilidad (y otras medidas) cuantifican la fuerza de la asociación en diferentes escalas, y nos referimos a ellas como medidas de asociación.

- $Pr[Y = 1|A = 1] - Pr[Y = 1|A = 0]$ (Diferencia de riesgo asociativo)
- $\frac{Pr[Y=1|A=1]}{Pr[Y=1|A=0]}$ (Razón de riesgo asociativo)
- $\frac{Pr[Y=1|A=1]/Pr[Y=0|A=1]}{Pr[Y=1|A=0]/Pr[Y=0|A=1]}$ (Odds ratio de asociación)

Si $Pr[Y = 1|A = 1] = Pr[Y = 1|A = 0]$, entonces $A \perp\!\!\!\perp Y$ (es decir, A, Y son independientes).

Causalidad Las inferencias sobre la causalidad están relacionadas con las preguntas en mundos contrafactuales, como "¿cuál sería el riesgo si todos hubieran sido tratados?" y "¿cuál sería el riesgo si no se hubiera tratado a todos?"

La diferencia entre asociación y causalidad: $Pr[Y = 1|A = 1]$ es una probabilidad condicional, mientras que, $Pr[Y^a = 1]$ es una probabilidad incondicional.

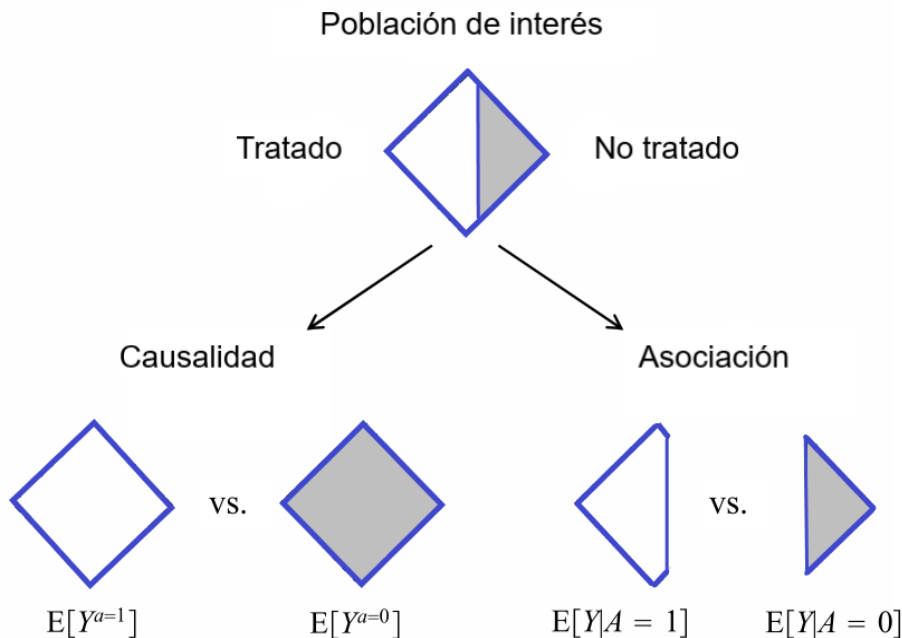
EJEMPLO

$Pr[Y^{a=1} = 1] = \frac{10}{20} = Pr[Y^{a=0} = 1]$. Por lo tanto, no hubo causalidad.

Las definiciones radicalmente diferentes entre asociación versus causalidad explican la frase "asociación no es causalidad". En nuestra población, hubo asociación porque el riesgo de mortalidad en los tratados (7/13) fue mayor que en los no tratados (3/7). Sin embargo, no hubo causalidad porque el riesgo si todos hubieran sido tratados (10/20) era el mismo que el riesgo si todos no hubieran sido tratados.

La Figura 1 representa la diferencia de asociación de causalidad. La población (representada por un diamante) se divide en un área blanca (la tratada) y un área gris más pequeña (la no tratada).

Figura 1: Diferencia de asociación-causalidad (Figura 1.1 en Hernán & Robins).



La definición de causalidad implica un contraste entre todo el diamante blanco (todos los individuos tratados) y todo el diamante gris (todos los individuos no tratados), mientras que la asociación implica un contraste entre las áreas blancas (tratadas) y grises (no tratadas) de Diamante original. La inferencia causal requiere datos como los datos hipotéticos en la Tabla 1, pero todo lo que tenemos son datos del mundo real como los de la Tabla 2. Una pregunta decisiva es entonces averiguar en qué condiciones se pueden usar los datos del mundo real para la inferencia causal. Una respuesta a esta pregunta es: realizar un experimento aleatorio.

2.6. EXPERIMENTOS ALEATORIOS

Un ejemplo de una pregunta causal: ¿Mirar hacia el cielo hace que otros peatones también miren hacia arriba? Los componentes principales de esta pregunta causal son: queremos saber si una acción A (su mirada hacia arriba) afecta un resultado Y (la mirada de otras personas hacia arriba) en una población específica (por ejemplo, residentes de Barcelona en 2020).

Supongamos que el reto es diseñar un estudio científico para responder esta pregunta. Después de pensar un poco, “puedo pararme en la acera y lanzar una moneda cada vez que alguien se acerca”. Si sale cara, voy a mirar hacia arriba; si sale cruz, miraré al frente. Repito el experimento unos miles de veces. Por lo tanto, nuestra solución al desafío fue realizar un experimento aleatorio. Fue un experimento porque el investigador (usted) llevó a cabo la acción de interés (mirar hacia arriba), y fue aleatorizado porque la decisión de actuar sobre cualquier sujeto de estudio (peatón) fue tomada por un dispositivo aleatorio (lanzamiento de monedas).

A continuación, presentamos métodos de diseño para experimentos aleatorios: Experimentos aleatorios marginales y experimentos aleatorios condicionales.

Diseño 1: Experimentos aleatorios marginales

Nos referimos a los experimentos marginalmente aleatorios cuando usamos una probabilidad de aleatorización incondicional (marginal) que es común a todos los individuos. En otras palabras, según este diseño, usaríamos una sola moneda para asignar tratamiento a todas las personas (por ejemplo, tratadas si son cruces, sin tratar si son caras). Se espera que un experimento aleatorio marginal produzca intercambiabilidad entre los grupos tratados y no tratados, denotado por $Y^a \perp\!\!\!\perp A$.

Existe intercambiabilidad incondicional (o marginal): $Y^a \perp\!\!\!\perp A$ sí y solo si

$$\Pr[Y^{a=i} = 1|A = 1] = \Pr[Y^{a=i} = 1|A = 0], i \in \{0,1\}$$

Diseño 2: Experimentos aleatorios condicionales

Nos referimos por experimentos aleatorios condicionales cuando usamos varias probabilidades de aleatorización que dependen de los valores de la variable L , es decir, nuestro experimento aleatorio condicional es simplemente la combinación de dos experimentos aleatorios marginales mutuamente excluyentes: uno realizado en el subconjunto de individuos en condición crítica ($L = 1$), el otro en el subconjunto de individuos en condición no crítica ($L = 0$).

Por lo tanto, bajo este diseño no usaríamos una sola moneda sino dos monedas para asignar tratamientos a todas las personas (por ejemplo, una moneda con probabilidad de **0.75** de girar las cruces para individuos en condiciones críticas $L = 1$, y otra moneda con una probabilidad de **0.5** de girar las cruces para individuos en condiciones no críticas $L = 0$).

Un experimento condicionalmente aleatorizado generalmente no dará como resultado la intercambiabilidad de los tratados y los no tratados porque, por diseño, cada grupo puede tener una proporción diferente de individuos con mal pronóstico, es decir, la intercambiabilidad $Y^a \perp\!\!\!\perp A$ no se cumple.

Aunque la aleatorización condicional no garantiza la intercambiabilidad incondicional (o marginal) $Y^a \perp\!\!\!\perp A$, garantiza la intercambiabilidad condicional $Y^a \perp\!\!\!\perp A|L$ dentro de los niveles de la variable L , es decir,

Existe intercambiabilidad condicional: $Y^a \perp\!\!\!\perp A|L$ sí y solo si

$$\Pr[Y^{a=i} = 1|A = 1, L = l] = \Pr[Y^{a=i} = 1|A = 0, L = l] \quad i, l \in \{0,1\}$$

2.6.1. Aleatorización

La aleatorización es un procedimiento que consiste en asignar aleatoriamente a los individuos un tratamiento A (1: tratado, 0: no tratado), garantizando que los valores faltantes, por ejemplo, de la Tabla 3 ocurrieron por azar.

Tabla 3: Valores faltantes en el resultado contrafactual (Tabla 2.1 en Hernán & Robins).

	A	Y	Y ⁰	Y ¹		A	Y	Y ⁰	Y ¹
Rheia	0	0	0	?	Leto	0	0	0	?
Kronos	0	1	1	?	Ares	1	1	?	1
Demeter	0	0	0	?	Athena	1	1	?	1
Hades	0	0	0	?	Hephaestus	1	1	?	1
Hestia	1	0	?	0	Aphrodite	1	1	?	1
Poseidon	1	0	?	0	Cyclope	1	1	?	1
Hera	1	0	?	0	Persephone	1	1	?	1
Zeus	1	1	?	1	Hermes	1	0	?	0
Artemis	0	1	1	?	Hebe	1	0	?	0
Apollo	0	1	1	?	Dionysus	1	0	?	0

Usamos la aleatorización como una herramienta que sirve para decidir quien recibe el tratamiento A , evitando los sesgos de selección. Un objetivo que se pretende alcanzar es hacer posible las comparaciones entre los grupos bajo tratamiento A , es decir, la aleatorización produce intercambiabilidad incondicional $Y^a \perp\!\!\!\perp A$ o intercambiabilidad condicional $Y^a \perp\!\!\!\perp A | L$, para experimentos aleatorios marginales y condicionales, respectivamente.

- Existe intercambiabilidad incondicional: $Y^a \perp\!\!\!\perp A$, sí y solo si

$$\Pr[Y^{a=i} = 1 | A = 1] = \Pr[Y^{a=i} = 1 | A = 0], i \in \{0,1\}$$

- Existe intercambiabilidad condicional: $Y^a \perp\!\!\!\perp A | L$, sí y solo si

$$\Pr[Y^{a=i} = 1 | A = 1, L = l] = \Pr[Y^{a=i} = 1 | A = 0, L = l] \quad i, l \in \{0,1\}$$

Importante:

- El cálculo de CRR es imposible debido a valores faltantes
- Se espera que la aleatorización produzca intercambiabilidad
- La intercambiabilidad significa que el resultado sería el mismo en ambos grupos de estudio bajo el supuesto de que todos los individuos fueron tratados o todos los individuos no fueron tratados.
- En experimentos aleatorizados ideales, la asociación es causalidad

2.6.2. Aleatorización condicional

La Tabla 4 muestra los datos hipotéticos de estudio aleatorizado de trasplante de corazón. Además de los datos sobre el tratamiento A (1 si el individuo recibió un trasplante, 0 en caso contrario) y el resultado Y (1 si el individuo murió, 0 en caso contrario), la Tabla 4 también contiene datos sobre el factor pronóstico L (1 si el individuo estaba en estado crítico, 0 en caso contrario), que fue medido antes de asignar el tratamiento.

Tabla 4 Ejemplo del “mundo real” sobre el estudio de trasplante de corazón, que incluye una tercera variable (Tabla 2.2 en Hernán & Robins)

	L	A	Y		L	A	Y
Rheia	0	0	0	Leto	1	0	0
Kronos	0	0	1	Ares	1	1	1
Demeter	0	0	0	Athena	1	1	1
Hades	0	0	0	Hephaestus	1	1	1
Hestia	0	1	0	Aphrodite	1	1	1
Poseidon	0	1	0	Cyclope	1	1	1
Hera	0	1	0	Persephone	1	1	1
Zeus	0	1	1	Hermes	1	1	0
Artemis	1	0	1	Hebe	1	1	0
Apollo	1	0	1	Dionysus	1	1	0

Aleatorización condicional son experimentos aleatorios con más de una moneda. Dado dos diseños de estudio mutuamente excluyentes estamos interesado en averiguar si la aleatorización garantiza la intercambiabilidad condicional $Y^a \perp\!\!\!\perp A \mid L$ dentro de los niveles de la variable L . Concretamente, necesitamos comprobar que los datos en la Tabla 4 han surgido de un experimento aleatorio condicional.

Por lo tanto, aunque la aleatorización condicional no garantiza la intercambiabilidad incondicional (o marginal) $Y^a \perp\!\!\!\perp A$, garantiza la intercambiabilidad condicional $Y^a \perp\!\!\!\perp A \mid L$ dentro de los niveles de la variable L . En las siguientes dos secciones describimos cómo usar los datos de experimentos aleatorios condicionales para calcular el efecto causal promedio en toda la población.

- En el ejemplo del “mundo real” los grupos de tratamiento no son intercambiables, ya que,

$$Pr[L = 1|A = 0] = \frac{3}{7}, \quad Pr[L = 1|A = 1] = \frac{9}{13}$$

- En el ejemplo del “mundo real”, la asignación del tratamiento fue condicional aleatorizado con probabilidades:

$$Pr[A = 1|L = 0] = 0.5, \quad Pr[A = 1|L = 1] = 0.75$$

Entonces, intercambiabilidad condicional, $Y^a \perp\!\!\!\perp A \mid L$, se mantiene dentro de los niveles del factor pronóstico L .

¿Cómo se puede calcular en un experimento condicionalmente aleatorizado? Una respuesta a esta pregunta es la estandarización o ponderación de la probabilidad inversa.

2.6.3. Estandarización

Teniendo en cuenta que la intercambiabilidad significa que el resultado sería el mismo en ambos grupos de estudio si ambos recibieron o si ambos no recibieron el tratamiento, la estandarización sirve para calcular la razón de riesgo causal en un experimento aleatorio condicional. Más formalmente, el riesgo contrafactual marginal bajo tratamiento $\Pr[Y^a = 1]$ es el promedio ponderado de los riesgos específicos del estrato $\Pr[Y^a = 1|L = 0]$ y $\Pr[Y^a = 1|L = 1]$ con pesos iguales a la proporción de individuos en la población con $L = 0$ y $L = 1$, respectivamente. Es decir,

$$\Pr[Y^a = 1] = \Pr[Y^a = 1|L = 0]\Pr[L = 0] + \Pr[Y^a = 1|L = 1]\Pr[L = 1]$$

Usando una notación más compacta,

$$\Pr[Y^a = 1] = \sum_l \Pr[Y^a = 1|L = l] \Pr[L = l]$$

Por intercambiabilidad condicional, el riesgo contrafactual es igual al riesgo observado, es decir,

$$\underbrace{\Pr[Y^a = 1|L = l]}_{\text{No observable}} = \underbrace{\Pr[Y = 1|L = l, A = a]}_{\text{Observable}}$$

Reemplazando en la expresión citada como una notación más compacta, obtenemos

$$\underbrace{\Pr[Y^a = 1]}_{\text{No observable}} = \underbrace{\sum_l \Pr[Y = 1|L = l, A = a] \Pr[L = l]}_{\text{Observable}}$$

Los riesgos estandarizados en los tratados y no tratados son iguales a los riesgos contrafactuales bajo tratamiento y sin tratamiento, respectivamente. Por lo tanto, la razón de riesgo causal $\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}$, puede calcularse mediante la estandarización como

$$\frac{\underbrace{\Pr[Y^{a=1}=1]}_{\text{No observable}}}{\underbrace{\Pr[Y^{a=0}=1]}_{\text{No observable}}} = \frac{\sum_l \Pr[Y = 1|L = l, A = 1] \Pr[L = l]}{\sum_l \Pr[Y = 1|L = l, A = 0] \Pr[L = l]}$$

El estudio de trasplante de corazón de la Tabla 4, es un experimento aleatorio condicional. El riesgo causal mediante estandarización se calcula como:

$$\text{CRR} = \frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]} = \frac{\Pr[Y=1|L=0, A=1]\Pr[L=0] + \Pr[Y=1|L=1, A=1]\Pr[L=1]}{\Pr[Y=1|L=0, A=0]\Pr[L=0] + \Pr[Y=1|L=1, A=0]\Pr[L=1]}$$

$\Pr[L = 0] = \frac{8}{20} = 0.4$, es decir, el 40% de los sujetos (8) están en el grupo $L = 0$.

$\Pr[L = 1] = \frac{12}{20} = 0.6$, es decir, el 60% de los sujetos (12) están en el grupo $L = 1$.

- Los riesgos observados son iguales a los riesgos contrafactuales en el grupo de 8 individuos en condición no crítica $L=0$. Es decir,

$$\Pr[Y = 1 | L = 0, A = 0] = \Pr[Y^{a=0} = 1 | L = 0],$$

$$\Pr[Y = 1 | L = 0, A = 1] = \Pr[Y^{a=1} = 1 | L = 0],$$

El riesgo de muerte entre los sujetos que no recibieron el trasplante de corazón es:

$$\Pr[Y = 1 | L = 0, A = 0] = \Pr[Y^{a=0} = 1 | L = 0] = 1/4$$

El riesgo de muerte entre los sujetos que recibieron el trasplante de corazón es:

$$\Pr[Y = 1 | L = 0, A = 1] = \Pr[Y^{a=1} = 1 | L = 0] = 1/4$$

- Los riesgos observados son iguales a los riesgos contrafactuales en el grupo de 12 sujetos en condición crítica $L=1$, es decir,

$$\Pr[Y = 1 | L = 1, A = 0] = \Pr[Y^{a=0} = 1 | L = 1]$$

$$\Pr[Y = 1 | L = 1, A = 1] = \Pr[Y^{a=1} = 1 | L = 1]$$

Entonces, el riesgo de muerte entre los sujetos que no recibieron el trasplante de corazón es:

$$\Pr[Y = 1 | L = 1, A = 0] = \Pr[Y^{a=0} = 1 | L = 1] = 2/3$$

El riesgo de muerte entre los sujetos que recibieron el trasplante de corazón es:

$$\Pr[Y = 1 | L = 1, A = 1] = \Pr[Y^{a=1} = 1 | L = 1] = 2/3$$

- Riesgo de muerte si todos los sujetos hubieran recibido el trasplante de corazón, $\Pr[Y^{a=1} = 1]$, es el promedio ponderado:

$$\Pr[Y^{a=1} = 1] = 1/4 \times 0.4 + 2/3 \times 0.6 = 0.5$$

- Riesgo de muerte si todos los sujetos no hubieran recibido el trasplante de corazón, $\Pr[Y^{a=0} = 1]$, es el promedio ponderado:

$$\Pr[Y^{a=0} = 1] = 1/4 \times 0.4 + 2/3 \times 0.6 = 0.5$$

Por lo tanto, la razón de riesgo causal es:

$$CRR = \frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]} = \frac{\Pr[Y=1|L=0,A=1]\Pr[L=0] + \Pr[Y=1|L=1,A=1]\Pr[L=1]}{\Pr[Y=1|L=0,A=0]\Pr[L=0] + \Pr[Y=1|L=1,A=0]\Pr[L=1]} = \frac{0.5}{0.5} = 1$$

2.6.4. Probabilidad inversa de ponderación

Modeliza la probabilidad de pertenencia de los sujetos al grupo de tratamiento al que pertenecen condicionado a los factores de confusión, y se ponderará por el inverso de esa probabilidad. Así los sujetos que tienen una probabilidad menor de pertenencia en ese grupo tendrán un peso mayor, y los sujetos que tienen una probabilidad mayor de pertenencia en ese grupo tendrán un peso menor. El objetivo de la ponderación de IP es crear una pseudopoblación en la que no exista asociación entre las covariables L y el tratamiento A . Los pesos de IP, denotado por $W^A = 1/f(A|L)$, simulan una pseudopoblación en la que todos los miembros de la población de estudio son reemplazados por dos copias de sí mismos. Una copia recibe el valor de tratamiento $A = 1$ y la otra copia recibe el valor de tratamiento $A = 0$.

EJEMPLO

Usando los datos de la Tabla 5, crear una pseudopoblación y computar la razón de riesgo causal mediante la ponderación de la probabilidad inversa.

Tabla 5 Ejemplo del “mundo real” con una tercera variable (Tabla 2.2 en Hernán & Robins).

	L	A	Y		L	A	Y
Rheia	0	0	0	Leto	1	0	0
Kronos	0	0	1	Ares	1	1	1
Demeter	0	0	0	Athena	1	1	1
Hades	0	0	0	Hephaestus	1	1	1
Hestia	0	1	0	Aphrodite	1	1	1
Poseidon	0	1	0	Cyclope	1	1	1
Hera	0	1	0	Persephone	1	1	1
Zeus	0	1	1	Hermes	1	1	0
Artemis	1	0	1	Hebe	1	1	0
Apollo	1	0	1	Dionysus	1	1	0

La idea es crear una pseudopoblación con valores $Y^{a=1}$ y $Y^{a=0}$ para todos los sujetos y el cómputo posterior de la CRR . Para este propósito, las observaciones reciben pesos que dependen de los valores de A y L . Para el caso discreto, los pesos se calculan con la siguiente fórmula:

$$W^A = 1/f(A|L) = 1/\Pr(A|L)$$

Como primer paso, es calcular el número de sujetos en condición no crítica ($L = 0$) y el número de sujetos en condición crítica ($L = 1$). Realizando un conteo rápido a partir de la Tabla 5, se obtiene resultados para la representación de la población original (datos de la Tabla 5) mediante la construcción de un diagrama de árbol.

ESCENARIO 1: Sujetos en condición no crítica ($L = 0$)

8 sujetos de 20 en condición no crítica ($L = 0$): $\Pr[L = 0] = 8/20 = 0.4$

De los 8 sujetos en condición crítica ($L = 0$)

4 sujetos no fueron tratados ($A = 0$): $\Pr[A = 0|L = 0] = 4/8 = 0.5$

4 sujetos fueron tratados ($A = 1$): $\Pr[A = 1|L = 0] = 4/8 = 0.5$

De los 4 sujetos no tratados ($A = 0$)

3 sujetos sobrevivieron ($Y = 0$): $\Pr[Y = 0|L = 0, A = 0] = 3/4 = 0.75$

1 sujeto murió ($Y = 1$): $\Pr[Y = 1|L = 0, A = 0] = 1/4 = 0.25$

Reciben un peso: $W^{A=0} = 1/\Pr(A = 0|L = 0) = 1/0.5 = 2$

De los 4 sujetos tratados ($A = 1$)

sujetos sobrevivieron ($Y = 0$): $\Pr[Y = 0|L = 0, A = 1] = 3/4 = 0.75$

1 sujeto murió ($Y = 1$): $\Pr[Y = 1|L = 0, A = 1] = 1/4 = 0.25$

Reciben un peso: $W^{A=1} = 1/\Pr(A = 1|L = 0) = 1/0.5 = 2$

ESCENARIO 2: Sujetos en condición crítica ($L = 1$)

12 sujetos de 20 en condición crítica ($L = 1$): $\Pr[L = 1] = 12/20 = 0.6$

De los 12 sujetos en condición crítica ($L = 1$)

3 sujetos no fueron tratados ($A = 0$): $\Pr[A = 0|L = 1] = 3/12 = 0.25$

9 sujetos fueron tratados ($A = 1$): $\Pr[A = 1|L = 1] = 9/12 = 0.75$

De los 3 sujetos no tratados ($A = 0$)

1 sujeto sobrevivió ($Y = 0$): $\Pr[Y = 0|L = 1, A = 0] = 1/3$

2 sujetos murieron ($Y = 1$): $\Pr[Y = 1|L = 1, A = 0] = 2/3$

Reciben un peso: $W^{A=0} = 1/\Pr(A = 0|L = 1) = 1/0.25 = 4$

De los 9 sujetos tratados ($A = 1$)

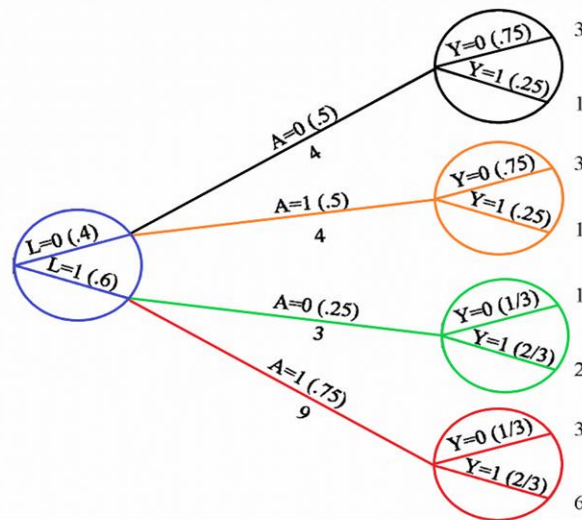
3 sujetos sobrevivieron ($Y = 0$): $\Pr[Y = 0|L = 1, A = 1] = 3/9 = 1/3$

6 sujetos murieron ($Y = 1$): $\Pr[Y = 1|L = 1, A = 1] = 6/9 = 2/3$

Reciben un peso: $W^{A=1} = 1/\Pr(A = 1|L = 1) = 1/0.75 = 1.33$

En la Figura 2 se presenta la construcción de un diagrama de árbol, con los resultados obtenidos del escenario 1 y 2.

Figura 2: Diagrama de árbol para la representación gráfica de la población original en el “ejemplo del mundo real” (Figura 2.1 en Hernán & Robins).



A partir de la Figura 2 simulamos la ponderación de IP para crear la pseudopoblación con valores $Y^{a=1}$ y $Y^{a=0}$.

Primero: Simulación de la ponderación por IP (si todos los sujetos son no tratados)

$$W^{A=0} = 1/f(A = 0|L) = 1/\Pr(A = 0|L) \Rightarrow \begin{cases} 1/\Pr(A = 0|L = 0) = 2 \\ 1/\Pr(A = 0|L = 1) = 4 \end{cases}$$

Observaciones en condición no crítica ($L = 0$)

3 sujetos sobrevivieron ($Y = 0$): ponderando por IP: $3 \times 2 = 6$

1 sujeto murió ($Y = 1$): ponderando por IP: $1 \times 2 = 2$

Observaciones en condición crítica ($L = 1$)

1 sujeto sobrevivió ($Y = 0$): ponderando por IP: $1 \times 4 = 4$

2 sujetos murieron ($Y = 1$): ponderando por su IP: $2 \times 4 = 8$

Segundo: Simulación de la ponderación de IP (si todos los sujetos son tratados)

$$W^{A=1} = 1/f(A = 1|L) = 1/\Pr(A = 1|L) \Rightarrow \begin{cases} 1/\Pr(A = 1|L = 0) = 2 \\ 1/\Pr(A = 1|L = 1) = 4/3 \end{cases}$$

Observaciones en condición no crítica ($L = 0$)

3 sujetos sobrevivieron ($Y = 0$): ponderando por IP: $3 \times 2 = 6$

1 sujeto murió ($Y = 1$): ponderando por IP: $1 \times 2 = 2$

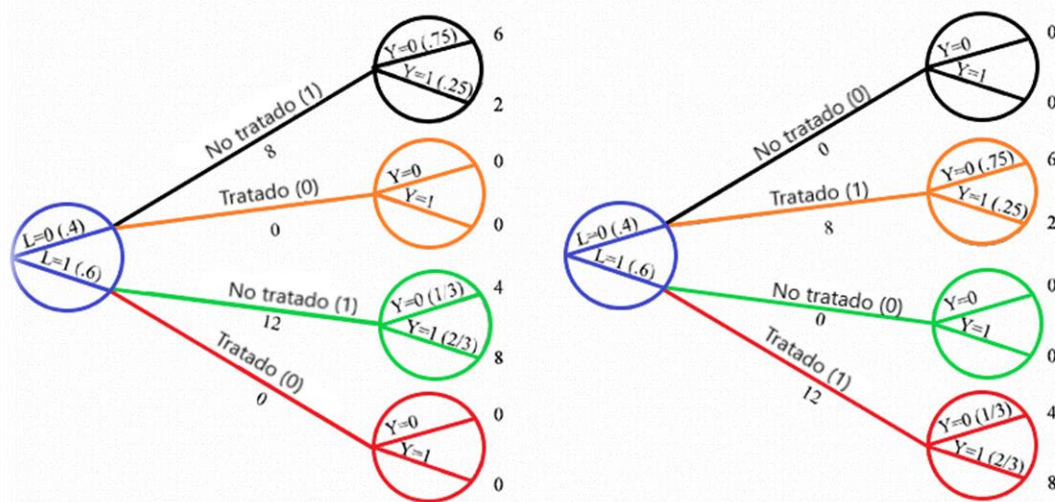
Observaciones en condición crítica ($L = 1$)

3 sujetos sobrevivieron ($Y = 0$): ponderando por IP: $3 \times 4/3 = 4$

6 sujetos murieron ($Y = 1$): ponderando por IP: $6 \times 4/3 = 8$

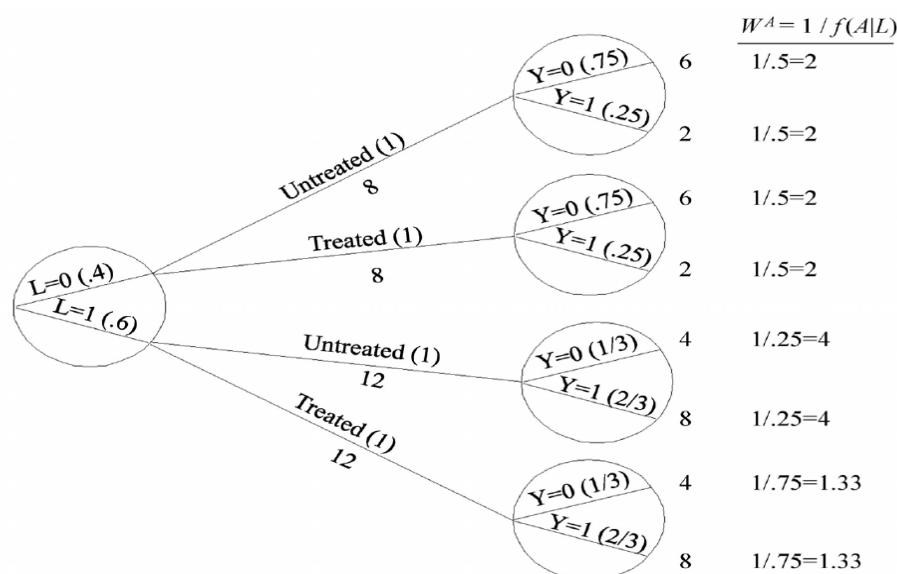
A continuación, se presenta dos diagramas de árbol de la simulación IP realizado hipotetizando primero, que todos los sujetos son no tratados y en segundo lugar hipotetizando que todos los sujetos son tratados.

Figura 3: Diagramas de árbol para la representación gráfica de la simulación, ponderación de la probabilidad inversa en el “ejemplo del mundo real”. Panel izquierdo: todos los sujetos son no tratados; panel derecho: todos los sujetos son tratados (Figura 2.2 en Hernán & Robins).



Ambas simulaciones se pueden unir para crear una población hipotética en la que cada individuo aparece como un individuo tratado y no tratado. Por lo tanto, uniendo ambas simulaciones, la población de estudio original de la **Figura 2** se transforma en la pseudopoblación de la **Figura 4**.

Figura 4: Representación gráfica de la pseudopoblación (Figura 2.3 en Hernán & Robins).



Es importante enfatizar que la pseudopoblación se crea al ponderar a cada individuo en la población por la inversa de la probabilidad condicional de recibir el nivel de tratamiento que efectivamente recibió. Esta idea básica, es el método ampliamente conocido como ponderación de la probabilidad inversa (IP). Tabla 6 resume los cálculos realizados, para la generación de la pseudopoblación..

Tabla 6 Ponderación de la probabilidad inversa y composición de la Pseudopoblación

L	A	Y	N Población Observada	$Pr[A L]$	W^A	N Pseudo Población
0	0	0	3	4/8	2	6
0	0	1	1	4/8	2	2
0	1	0	3	4/8	2	6
0	1	1	1	4/8	2	2
1	0	0	1	3/12	4	4
1	0	1	2	3/12	4	8
1	1	0	3	9/12	4/3	4
1	1	1	6	9/12	4/3	8

De la pseudopoblación, la razón de riesgo causal **CRR**, se calcula como:

$$CRR = \frac{Pr[Y^{a=1} = 1]}{Pr[Y^{a=0} = 1]} = \frac{(2 + 8)/20}{(2 + 8)/20} = 1$$

Bajo intercambiabilidad condicional $Y^a \perp\!\!\!\perp A \mid L$ en la población original, los tratados y los no tratados son (incondicionalmente) intercambiables en la pseudopoblación porque L es independiente de A . Es decir, la razón de riesgo asociativo en la pseudopoblación es igual a la razón de riesgo causal tanto en la pseudopoblación como en la población original" (Hernán & Robins, 2020).

2.7. ESTUDIOS OBSERVACIONALES

1965, Cochran, definió un estudio observacional como una investigación empírica en la que "el objetivo es dilucidar las relaciones de causa y efecto... [En entornos en los que] no es factible utilizar la experimentación controlada, en el sentido de poder imponer los procedimientos o tratamientos cuyos efectos se desean descubrir, o asignar a sujetos al azar a diferentes procedimientos".

Cochran tiene razón, muchas veces realizar un ensayo aleatorio no será ni ético, ni factible. Por ejemplo, un estudio sobre trasplante de corazón, es muy probable que un comité ético no apruebe. Los corazones escasean y, además, la sociedad favorece que se asignara a individuos que tienen más probabilidades de beneficiarse del trasplante, en lugar de una asignación aleatoria a los posibles receptores. Incluso si ignoramos los problemas éticos, la asignación aleatoria es imposible, puede que no haya corazones compatibles para aquellos sujetos asignados aleatoriamente al trasplante. Incluso si el estudio es factible, aún nos tomaría varios años completarlo, y no sería práctico hacer el estudio.

Entonces, las mejores inferencias causales deben proceder de estudios observacionales de alta calidad. Además, muchos estudios científicos no son experimentales. Por ejemplo, estudios sobre la evolución, las placas tectónicas, el calentamiento global o la astrofísica. La gran parte del conocimiento humano se deriva de estudios observacionales.

A continuación, revisamos bajo qué condiciones los estudios observacionales, son suficientemente fuertes para proporcionar evidencias causales firmes. Las inferencias causales sobre datos observacionales gira en torno del paradigma de experimentos aleatorizados (Intercambiabilidad, Positividad, Intervenciones bien definidas, Causalidad o predicción). Las condiciones bajo las cuales los estudios observacionales conducen a inferencias causales válidas, se han denominado, como las condiciones de identificabilidad.

2.7.1. Condiciones de identificabilidad

Se deben cumplir tres condiciones o supuestos para que un estudio observacional pueda ser conceptualizado como un experimento condicionalmente aleatorizado:

1. Tratamientos muy bien definidos (un requisito previo para la inferencia causal)
2. La probabilidad condicional de recibir el tratamiento depende de las covariables observadas (idea de intercambiabilidad)
3. La probabilidad condicional de recibir el tratamiento debe ser positiva (idea de Positividad).

Bajo estos supuestos, es decir, si se cumplen estas tres condiciones (de identificabilidad),

“... Los efectos causales se pueden identificar a partir de estudios observacionales utilizando estandarización o por ponderación de IP”

En cambio, para identificar los efectos causales en experimentos aleatorizados ideales no hace falta verificar línea por línea estas tres condiciones de identificabilidad, ya que, estos supuestos se cumplen por diseño. Para las inferencias causales solo hace falta disponer de los datos experimentales, mientras para los estudios observacionales necesitamos los datos de observación y condiciones de identificabilidad.

2.7.2. Intercambiabilidad

En experimentos incondicionalmente aleatorizados, los sujetos tratados y no tratados son intercambiables $Y^a \perp\!\!\!\perp A$, en otras palabras, significa que, si todos los sujetos tratados hubieran permanecido sin tratamiento, habrían experimentado el mismo resultado promedio que los sujetos no tratados, y si todos los sujetos no tratados hubieran permanecido con tratamiento, habrían experimentado el mismo resultado promedio que los sujetos tratados.

Esto es así porque la aleatorización asegura que los predictores (L_1, L_2, \dots) independientes del resultado Y se equilibren entre los grupos tratados y no tratados. Por lo tanto, en experimentos marginalmente aleatorios no hay presencia de desequilibrios en la distribución de los predictores L entre los grupos tratados y no tratados.

En cambio, en experimentos condicionalmente aleatorios, es de esperar la presencia de desequilibrio en la distribución de los predictores L entre los grupos tratados y los no tratados, ya que, por diseño, la probabilidad de recibir tratamiento depende de L . Sin embargo, los grupos tratados y los no tratados son intercambiables condicionalmente dentro de los niveles de la variable L . Por lo tanto, intercambiabilidad condicional $Y^a \perp\!\!\!\perp A \mid L$ se cumple en experimentos condicionalmente aleatorios porque, dentro de los niveles de L , todos los demás predictores del resultado se distribuyen por igual entre los grupos tratados y no tratados.

Hasta aquí, no hemos dicho nada acerca de la intercambiabilidad en los estudios observacionales. Es bien conocido, que cuando los investigadores no asignan el tratamiento al azar, es probable que las razones para recibir el tratamiento estén asociadas con algunos predictores (L_1, L_2, \dots) de resultados Y . Por ejemplo, si los médicos tienden a asignar los escasos trasplantes de corazón a quienes más lo necesitan, es decir, individuos en estado crítico $L = 1$. Entonces es de esperar, que la distribución de predictores de resultados varíe entre los grupos tratados y no tratados.

En los estudios observacionales no existe una fórmula analítica o empírica, para verificar la intercambiabilidad condicional $Y^a \perp\!\!\!\perp A \mid L$ en ausencia de aleatorización. Por ejemplo, la intercambiabilidad condicional aleatoria $Y^a \perp\!\!\!\perp A \mid L$ no se mantendrá si existen predictores independientes no medidos U del resultado, de modo que la probabilidad de recibir tratamiento A depende de U dentro de los estratos de L .

Incluso si la intercambiabilidad condicional aleatoria $Y^a \perp\!\!\!\perp A \mid L$ logramos garantizar, no se puede verificar empíricamente que ese sea realmente el caso, porque no tenemos datos

medidos disponibles sobre U para verificar empíricamente que la distribución de U es igual en los grupos tratados y los no tratados. Al analizar un estudio observacional bajo intercambiabilidad condicional, debemos medir muchas variables relevantes L en lugar de una sola variable, para que la condición de intercambiabilidad sea al menos aproximadamente verdadera. Por lo tanto, la validez de las inferencias causales observacionales necesita que el conocimiento experto de los investigadores sea correcto.

2.7.3. Positividad

Comenzamos diciendo que, en los experimentos marginalmente aleatorios, las probabilidades $Pr[A = 1]$ y $Pr[A = 0]$ son positivas por diseño. También, en experimentos condicionalmente aleatorizados, las probabilidades condicionales $Pr[A = 1 | L = l]$ y $Pr[A = 0 | L = l]$ son positivas por diseño.

La positividad se mantiene, condicional a L , si ninguna de las probabilidades condicionales $Pr[A = a | L = l]$ es 0 ni 1 para todos los niveles de A en el contraste causal, es decir, existe positividad si $Pr[A = a | L = l] > 0$.

Si tuviéramos $Pr[A = 1 | L = l] = 0.75$ y $Pr[A = 0 | L = l] = 0.50$. Entonces, la positividad se mantiene, condicional en L , porque ninguna de estas probabilidades es 0 ni 1.

El riesgo estandarizado y ponderado de IP solo es significativo cuando se cumple la condición de positividad. Si tuviéramos $Pr[A = 1 | L = l] = 1$ implicaría que $Pr[A = 0 | L = l] = 0$. Por lo tanto, la positividad no es válida para L .

Cuando no se cumple la positividad, el riesgo estandarizado y ponderado de IP no está bien definido, es decir, no podemos calcular la razón de riesgo causal. En estudios observacionales, ni la positividad ni la intercambiabilidad están garantizadas, aunque, la positividad a veces puede verificarse empíricamente, Hernán y Robins (2020).

Además, en estudios observacionales, el valor de A probablemente depende de algunos predictores de resultados L_1, L_2, \dots , por otro lado, es imposible saber si son todos los predictores L_i con distribuciones desiguales entre los grupos de tratamiento observados, es decir, no hay garantía de sostener que $Y^a \perp\!\!\!\perp A | L$. También, si hay múltiples versiones de un tratamiento, las intervenciones son no bien definidas y los problemas debido a intervenciones no especificadas no pueden resolverse aplicando métodos estadísticos sofisticados para inferencias causales. En el siguiente capítulo veremos cómo la metodología del propensity score intenta hacer posible la inferencia causal, en los estudios observacionales.

3. PROPENSITY SCORE

En este capítulo vamos a discutir puntajes de propensión: cómo y por qué se usa en la investigación. ¿Qué forma de magia estadística puede tomar un estudio observacional y convertirlo en un ensayo aleatorio? Si bien algunos afirman que son casi tan buenos como un ensayo aleatorio, en esta inmersión profunda, explicamos cómo funcionan y cuando no. El problema que los puntajes de propensión intentan resolver es omnipresente en la investigación médica.

Desea saber si un tratamiento causa un cierto resultado, pero todo lo que tiene son datos observacionales. Uno de los inconvenientes de los ensayos objetivos, son las distribuciones de las covariables que difieren generalmente entre los grupos de tratamiento. Para corregir estos desequilibrios presentaremos cuatro métodos de vanguardia basados en propensity score y su posterior computo de los efectos causales.

Motivación

Como ejemplo, supongamos que disponemos de datos sobre el consumo de marihuana y digamos que queremos averiguar si el consumo de marihuana conduce a la depresión. Por otro lado, supongamos que los datos sin procesar muestran una asociación clara. Las personas que fuman marihuana tienen más probabilidades de deprimirse. Pero surge un grado de especulación ya que, tal vez personas con ciertos rasgos que conducen a la depresión podrían decidir usar marihuana.

La evidencia cruda podría ser lo suficientemente buena para algunos, pero no para nosotros. Dado al escepticismo, podemos identificar todo tipo de factores asociados con fumar marihuana, es decir, cosas que afectan su propensión a fumar marihuana. Por ejemplo: ser hombre, ser blanco, fumar tabaco, etc. Luego, podríamos ajustar mediante un modelo multivariable para todos esos factores. Esa es una práctica estándar, pero hay un costo. Cada variable que se ajusta hace que la estimación de efecto de la marihuana en la depresión sea menos precisa y los intervalos de confianza alrededor de su estimación se amplían.

Perry Wilson profesor de la universidad de Yale (USA): “Simulé 100 variables que están relacionadas con el uso de marihuana, pero no con la depresión. A medida que ajusto a más y más de esas variables, mi estimación del vínculo de la marihuana con la depresión se vuelve cada vez menos precisa. Y este problema se hace más grande cuanto más pequeño es el estudio. Eventualmente, los intervalos de confianza son tan amplios que no tienes idea de lo que está pasando, y te has disparado en el pie con tus estadísticas elegantes. Entonces en lugar de un ajuste multivariable, construyes un puntaje de propensión. Un puntaje de propensión es un número único que integra todos los factores asociados con la exposición. Nos permite tomar esas 100 variables y condensarlas en una sola variable que da la probabilidad de que seas fumador de marihuana. Podemos usar esa única variable para evitar pagar el precio estadístico de ajustar por 100 cosas. Ahora solo podemos ajustar por uno, o bien, podemos encontrar personas con puntajes de propensión similares y unirlos.”

Al unir un individuo que fuma marihuana con otro individuo que no fuma marihuana, pero ambos, con puntaje de propensión 30%, significa que, en promedio, los dos individuos comparten un riesgo general de fumar marihuana, es por eso que tienen el mismo puntaje de propensión, afirma el profesor Perry Wilson de la universidad de Yale (USA).

En otras palabras, es como si ambos individuos estuvieran en un ensayo aleatorizado y uno fuera asignado a fumar marihuana y mientras que el otro fuera asignado al grupo de no fumadores de marihuana. Entonces, podemos comparar sus resultados para obtener una estimación más fiable del verdadero efecto de la marihuana, en la depresión.

Otra cosa que señala el profesor Perry Wilson de la universidad de Yale (USA) es que el emparejamiento basado en puntaje de propensión es tan bueno como su capacidad de predecir si el individuo estará expuesto o no. Señala que, si el modelo predictivo es inútil, el emparejamiento por propensión también es inútil. Por otro lado, afirma que, si el modelo de predicción es demasiado bueno, el emparejamiento por propensión no funciona. Termina señalando que el propensity score es una técnica estadística que depende de buenos modelos, pero no de una gran construcción de modelos y que las técnicas de puntaje de propensión funcionan mejor que un ajuste multivariable, en ciertas situaciones raras.

3.1. DEFINICIÓN DEL PROPENSITY SCORE

Para finales de 1982 el concepto de propensity score no existe, los científicos luchan por encontrar un nuevo método para eliminar los desequilibrios en las distribuciones de las covariables entre los grupos de tratamiento y estimar efectos marginales. Finalmente, en 1983, se introduce por primera vez un nuevo método de vanguardia para controlar el sesgo en los estudios observacionales debido a la no asignación aleatoria del sujeto al tratamiento. Es un método revolucionario que consiste en el cálculo de la probabilidad individual de recibir el tratamiento de interés condicionada por una serie de covariables observadas conocida como propensity score (P.R., 1983).

Definición del propensity score El propensity score para cada sujeto i , denotado por PS_i , se define como la probabilidad condicional de asignar al tratamiento $A = 1$ dada su vector de covariable $X = (X_1, X_2, \dots, X_n)$. En símbolos, el modelo del propensity score es la siguiente expresión

$$PS_i = Pr(A = 1|X) = Pr(A = 1|X_1, X_2, \dots, X_n); \quad 0 \leq PS_i \leq 1$$

Es la probabilidad de tratamiento dada las covariables X_1, X_2, \dots, X_n , para cada sujeto i .

Más concretamente, los propensity scores, son números que varían entre cero y uno derivados de las probabilidades condicionales. La condición es el patrón de covariables que presenta cada sujeto de interés. Por ejemplo, en un estudio sobre la supervivencia asociada a un tratamiento del trasplante de corazón y solo interesasen tres covariables previas (factor pronóstico, edad, sexo), en esta situación, el propensity score para un sujeto i , se modela como:

$$PS_i = Pr(A = 1|\text{factor pronóstico, edad, sexo}); \quad 0 \leq PS_i \leq 1$$

En esta expresión PS_i se interpreta como la probabilidad de recibir el trasplante de corazón dadas las covariables (factor pronóstico, edad y sexo), para cada sujeto i .

Si el valor de PS_i está cerca de cero significa que el sujeto i tiene una baja probabilidad de recibir trasplante de corazón. Y si el PS_i está cerca de 1 quiere decir que el sujeto i tiene una alta probabilidad de recibir trasplante de corazón. Es decir, el propensity score PS_i mide la propensión de cada sujeto i a recibir tratamiento dadas sus covariables (factor pronóstico, edad y sexo).

3.1.1. Propiedades del propensity score

A continuación, presentamos alguna de las propiedades básicas del propensity score:

Propiedad 1: Los sujetos con el mismo propensity score PS_i generalmente tienen valores diferentes en algunas covariables X_1, X_2, \dots, X_n .

Propiedad 2: Si la distribución de PS es el mismo para los sujetos tratados $A = 1$ y no tratados $A = 0$, entonces no hay confusión debido a las covariables X_1, X_2, \dots, X_n .

Propiedad 3: El propensity score solo equilibra las covariables medidas X_1, X_2, \dots, X_n , no garantiza la confusión residual por factores no medidos.

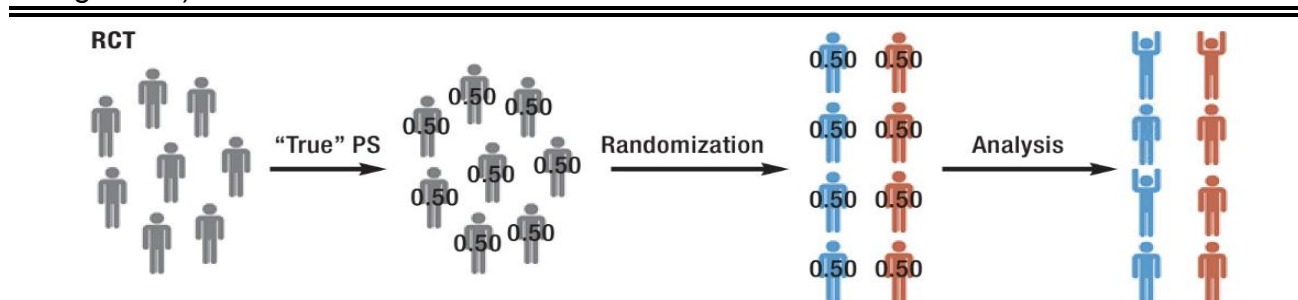
Propiedad 4: Si las covariables X_1, X_2, \dots, X_n son suficientes para ajustar el sesgo de confusión y selección, PS también es suficiente.

Propiedad 5: La intercambiabilidad de los tratados y los no tratados dentro de los niveles de las covariables X_1, X_2, \dots, X_n implica intercambiabilidad dentro de los niveles de puntaje de propensión PS , es decir, la intercambiabilidad condicional $Y^a \perp\!\!\!\perp A | X_1, X_2, \dots, X_n$ implica $Y^a \perp\!\!\!\perp A | PS$.

3.1.2. Estimación del propensity score

En un ensayo aleatorio ideal en el que la mitad de los sujetos son asignados al tratamiento $A = 1$, el propensity score (PS) es conocido y vale 0.5, es decir, $PS_i = 0.5$ para cada sujeto i . Además, $PS_i = 0.5$ para cualquier elección de covariables X_1, X_2, \dots, X_n , en otras palabras, el propensity score es invariante en ensayos aleatorios. En la Figura 5 se presenta un ensayo controlado aleatorio con el PS conocido.

Figura 5: Ilustración de la similitud entre un ensayo controlado aleatorio (ECA) y un análisis de puntaje de propensión (PS) (Sección inferior de la Figura en Kuss, Blettner & Börgerman).



Sujetos antes de someterse al tratamiento de interés (gris)

Sujetos tratados (azul) y sujetos no tratados (rojo)

Por el contrario, en los estudios observacionales, algunos sujetos pueden resultar ser más propensos a recibir tratamiento que otros, este hecho implica que la asignación del tratamiento esté fuera de control de los investigadores, en consecuencia, el valor verdadero del propensity score PS_i es desconocida para cada sujeto y depende de sus covariables X_1, X_2, \dots, X_n .

Por lo tanto, el PS_i solo podemos estimar a partir de los datos de observación utilizando el modelo de regresión logística, aunque también hay en la literatura otras metodologías que describiremos más adelante. Hemos dicho que en los estudios observacionales el propensity score PS_i solo se puede estimar, entonces surge una pregunta, qué covariables: X_1, X_2, \dots, X_n se incluyen en el modelo del propensity score. En la siguiente sección discutimos al respecto.

3.1.3. Selección del modelo del propensity score

Comenzamos enfatizando que en el modelo del propensity score solo se debe incluir variables que se miden al inicio y no covariables posteriores al inicio que pueden ser influenciadas o modificadas por el tratamiento A . En la práctica, identificar las covariables para incluir en el modelo del propensity, no es una tarea fácil, además, en la literatura aplicada no se ha consensuado sobre variables incluir en el modelo.

Según, Peter C. Austin (2011) hay cuatro posibles conjuntos de covariables para incluir en el modelo del propensity score.

$$PS_i = Pr(A = 1|X_1, X_2, \dots, X_n)$$

Modelo 1: Se obtiene al incluir todas las covariables medidas: X_1, X_2, \dots, X_n .

Modelo 2: Se obtiene al incluir todas las covariables X_1, X_2, \dots, X_n , que afectan la asignación del tratamiento A .

Modelo 3: Se obtiene al incluir todas las covariables: X_1, X_2, \dots, X_n , que afectan el resultado Y (es decir, los potenciales factores de confusión).

Modelo 4: Se obtiene al incluir todas las covariables X_1, X_2, \dots, X_n , que afectan a la asignación del tratamiento A y el resultado Y (es decir, los verdaderos factores de confusión).

Estos cuatro modelos del propensity score tienen sustento teórico riguroso, también, las evidencias empíricas lo sustentan. En 2007, Austin, Grootendorst y Anderson, examinaron el uso de cada uno de los cuatro conjuntos diferentes de covariables para incluir en el modelo del propensity score.

En un contexto general:

- Se demostró que había ventajas usando el modelo 3 y modelo 4.

En el contexto del propensity score matching “(se explicará más adelante)”

- El uso de cualquiera de los cuatro modelos equilibra las covariables entre los sujetos tratados y no tratados en la muestra emparejada.
- Cuando se usa el modelo 3 o modelo 4, no se obtiene el equilibrio entre los sujetos tratados y no tratados en la muestra emparejada en aquellas variables que son independientes del resultado Y , pero que afectan la asignación del tratamiento A . Sin embargo, se forman un mayor número de pares emparejados usando el modelo 3 o modelo 4, en comparación con el modelo 1 y modelo 2.
- El uso del modelo 3 y modelo 4 da estimaciones del efecto del tratamiento con mayor precisión.

Por otro lado, Brookhart et al. (2006), sugiere que las variables que no afectan el tratamiento A , pero si al resultado Y siempre deben incluirse en el modelo del propensity score. Además, enfatiza que incluir variables que afectan el tratamiento A , pero no al resultado Y aumenta la varianza del efecto del tratamiento estimado sin una reducción asociada al sesgo.

Breve resumen de la literatura publicada de (Peter C. Austin, 2011): Afirma, que en la práctica es difícil de clasificar con precisión las variables de referencia, los que solo afectan el resultado Y , los que solo afectan el tratamiento A y los que no afectan ni el tratamiento A ni el resultado Y . La mayoría de las variables a nivel sujeto probablemente afecten tanto la asignación del tratamiento como el resultado. Por lo tanto, es preferible, de manera segura, incluir todas las covariables medidas.

Las variables que denotan diferentes periodos temporales son las que requieren una mayor investigación. Por ejemplo, en un estudio que compara el efecto de un tratamiento más antiguo con el de un tratamiento más nuevo. Los sujetos que ingresaron al estudio en un periodo anterior pueden ser más propensos a recibir el tratamiento antiguo, mientras que los sujetos que ingresaron al estudio en un periodo posterior pueden ser más propensos a recibir el tratamiento nuevo. Por lo tanto, una variable que denota un periodo temporal afectaría la asignación del tratamiento. Sin embargo, si el resultado Y es condicionalmente independiente del periodo temporal, al incluir una variable que denota un periodo temporal en el modelo del propensity score, podría dar lugar a la formación de menos pares emparejados, es preferible no incluirla variables que denotan periodo temporal en el modelo del propensity score.

El objetivo de modelizar correctamente el propensity score, según, Hernán y Robins (2020) “... el propensity score basado en un modelo correcto proporciona un mejor equilibrio en general”. Después de modelizar el propensity score, una pregunta crucial es, cómo se obtiene el propensity score. Una respuesta a esta pregunta es, regresión logística.

3.1.4. Modelo de regresión logística

La regresión logística en el contexto multivariante es el modelo matemático ampliamente utilizado en la investigación biomédica. Usamos modelo de regresión logística para estudiar factores de riesgo entre una variable de resultado (Y) y una variable de tratamiento (A) de interés, en situaciones, cuando: A es una variable continua o existen varios posibles factores de confusión C_1, \dots, C_k o uno de los posibles factores de confusión es una variable continua. Una ventaja del modelo de regresión logística es su capacidad para controlar muchas variables simultáneamente. Aunque el ajuste multivariable es atractivo, también es problemática, ya que, al incluir demasiadas variables en un modelo, las estimaciones pueden resultar incorrectas.

A continuación, presentamos una formulación del modelo de regresión logística, basado en las transparencias de la asignatura de Klaus Langohr (Curso 2018/19: Epidemiología).

Formulación del modelo de regresión logística

Sea $X = (X_1, \dots, X_m)$ el vector de covariables del modelo, que incluye tanto la variable del tratamiento A como los posibles factores de confusión C_1, \dots, C_k . Sea Y el resultado de interés:

$$Y = \begin{cases} 1: \text{Resultado presente,} \\ 0: \text{Resultado ausent,} \end{cases}$$

Como Y es una variable binaria, un modelo lineal como:

$$Y = \alpha + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon$$

No tiene sentido. En cambio, la probabilidad de que el resultado esté presente $Y = 1$ condicionado al vector de variables X , se usa como la variable de respuesta en un modelo de regresión.

$$p = \Pr(Y = 1|X) = \Pr(Y = 1|X_1, \dots, X_m) \in (0, 1)$$

p : es la probabilidad de enfermar o riesgo de enfermar

Dado que $p \in (0, 1)$, la regresión logística modela la transformación **logit** de p , cuyo rango es el conjunto de los números reales \mathbb{R} , como la combinación lineal de las variables independientes X_1, \dots, X_m . El **logit** del modelo de regresión logística (es decir, predictor lineal) viene dado por la ecuación:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \dots + \beta_m X_m = \alpha + \beta'X$$

Esta expresión es equivalente a

$$p = \frac{\exp\{\alpha + \beta_1 X_1 + \dots + \beta_m X_m\}}{1 + \exp\{\alpha + \beta_1 X_1 + \dots + \beta_m X_m\}} = \frac{\exp(\alpha + \beta'X)}{1 + \exp(\alpha + \beta'X)}$$

- X_i es un factor de riesgo para Y , si $\beta_i > 0$ o $\beta_i < 0$.
- Y y X son independientes, si $\beta_i = 0$.

Basado en el modelo de regresión logística, a continuación, explicamos cómo se obtiene el propensity score PS .

3.1.5. Estimación del propensity score mediante regresión logística

El propensity score:

$$PS = Pr(A|X_1, X_2, \dots, X_n)$$

Se obtiene mediante un modelo de regresión logística, fijando la variable indicadora del tratamiento A como variable dependiente. Es decir, para estimar el puntaje de propensión PS se ajusta todas las covariables X_1, \dots, X_m en el modelo de regresión logística sin incluir el resultado.

$$Pr(A = 1|X_1, \dots, X_m) = \frac{\exp(\alpha + \beta'X)}{1 + \exp(\alpha + \beta'X)}$$

Del modelo ajustado, se obtienen los parámetros de la regresión: $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_n$.

El propensity score estimado PS , viene dada por:

$$PS = \frac{\exp(\hat{\alpha} + \hat{\beta}'X)}{1 + \exp(\hat{\alpha} + \hat{\beta}'X)}$$

El propensity score estimado, para cada sujeto i , se denota por PS_i

$$PS_i = \frac{\exp(\hat{\alpha} + \hat{\beta}'X_i)}{1 + \exp(\hat{\alpha} + \hat{\beta}'X_i)}$$

Los propensity scores oscilarán entre 0 y 1 y reflejarán la probabilidad de tratamiento basada en las covariables. El *logit* del propensity score PS es la siguiente expresión:

$$\text{logit}(PS) = \ln\left(\frac{PS}{1 - PS}\right) = \hat{\alpha} + \hat{\beta}'X$$

Cuando la variable indicadora del tratamiento A , tiene más de dos niveles, el propensity score PS se obtiene del modelo de regresión logística multinomial. Existen otros métodos de estimación para obtener el propensity score (PS).

3.1.6. Otros métodos para estimar el propensity score

En la literatura existen otros métodos para estimar PS, por ejemplo: Se puede estimar mediante modelos probit, análisis discriminante, clasificador Bayesiano, bosques aleatorios, redes neuronales, análisis de supervivencia y otros. Describiremos brevemente los cuatro primeros enfoques.

3.1.6.1. Modelos Probit

Si el estudio está basado en el campo econométrico es preferible usar regresión probit para la obtención del propensity score. Este método consiste, primero en ajustar el vector de covariables $X = (X_1, \dots, X_n)$ en el modelo probit.

$$Pr[A = 1|X] = \Phi(\alpha + \beta'X)$$

Para la estimación de los parámetros $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_n$ se ajusta el modelo en R a través de la función `glm()`.

Finalmente, el propensity score, para cada individuo i se obtiene evaluando $\hat{\alpha} + \hat{\beta}'X_i$ en la función de distribución acumulada normal estándar $\Phi(z)$.

$$PS_i = \Phi(\hat{\alpha} + \hat{\beta}'X_i) = \int_{-\infty}^{\hat{\alpha} + \hat{\beta}'X_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt$$

3.1.6.2. Análisis discriminante

El análisis discriminante, introducido por Fisher (1936), es un tecnicismo que se utiliza para predecir la pertenencia a un grupo de tratamiento a partir del vector de covariables $X = (X_1, \dots, X_n)$. Mediante este método el propensity score para una variable indicadora de tratamiento dicotómico, se obtiene ajustando el vector de covariables en la función discriminante

$$Pr[A = 1|X] = \alpha + \beta'X$$

Para la estimación de los parámetros $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_n$, se ajusta el modelo lineal discriminante en R, usando la función `lda()`. El propensity score estimado para cada individuo i , se obtiene como

$$PS_i = \hat{\alpha} + \hat{\beta}'X_i$$

3.1.6.3. Clasificador Bayesiano

Es una técnica de clasificación y predicción supervisada en el sentido de que las covariables X_1, \dots, X_n son categóricas o continuas. Esta técnica consiste en construir modelos que predicen la probabilidad de posibles resultados basado en el Teorema general de Bayes.

$$Pr[A | X_1, \dots, X_n] = Pr[A]Pr[X_1|A]Pr[X_2|A] \dots Pr[X_n|A]$$

$Pr[A]$: Es la probabilidad a priori del grupo de tratamiento y se computa generalmente considerando cada grupo equiprobable, es decir,

$$Pr[A = 1] = Pr[A = 0] = 0.5$$

$Pr[X_1|A], Pr[X_2|A], \dots, Pr[X_n|A]$: Para computar estas probabilidades condicionales se siguen los siguientes pasos,

Paso 1: Para cada grupo de tratamiento se crea una tabla y se realiza un conteo de los valores en cada categoría de las covariables.

Paso 2: Aplicar la corrección de Laplace, para evitar que las probabilidades condicionales sean diferentes de cero. De manera formal, aplicar corrección de Laplace significa sumar una unidad a cada valor del Paso 1.

Paso 3: Normalizar para obtener un rango de valores entre 0 y 1. Normalizar significa dividir cada valor de la celda por la suma de los valores de la fila.

Después de la descripción del cálculo de la probabilidad a priori y probabilidades condicionales, el propensity score para cada individuo i se obtiene de la siguiente manera:

$$PS_i = \max\{Pr[A = 1|X_i]; Pr[A = 0|X_i]\}$$

Es importante señalar que si el máximo argumento es $Pr[A = 1|X_i]$ al individuo i se asigna al grupo de tratamiento $A = 1$, con su respectivo valor del propensity score PS_i . Si el máximo argumento es $Pr[A = 0|X_i]$ al individuo i se asigna al grupo de tratamiento $A = 0$.

El propensity score mediante el clasificador Bayesiano en R se puede estimar mediante el uso de la función `naiveBayes()`.

3.1.6.4. Bosques aleatorios

Es uno de los métodos basados en árboles de clasificación surgido en el ámbito del aprendizaje automático (Machine Learning). Es una técnica que ya no se basa en modelos estadísticos, sino, en algoritmos de clasificación utilizando particiones iterativas. En la literatura, recomiendan obtener el propensity score mediante bosques aleatorios cuando los métodos convencionales de regresión fallan o se dispone de datos de alta complejidad. Una idea básica, para la obtención del propensity score mediante este método consiste, primero, en construir un árbol para cada sujeto en el bosque aleatorio, utilizando la variable indicadora del tratamiento A como salida y las covariables X_1, \dots, X_n como entradas.

Una vez que se construye un árbol, el propensity score se obtiene fácilmente, utilizando solo la información contenida en los nodos terminales. Después de la descripción de algunos tecnicismos de la estimación del propensity score, una pregunta es sobre el uso del propensity score. En el siguiente capítulo se presenta el uso del propensity score mediante cuatro metodologías de vanguardia.

3.2. METODOS DEL PROPENSITY SCORE

En los ensayos objetivos, existen cuatro métodos basados en propensity score para corregir los desequilibrios entre los grupos de tratamiento debido a la no aleatorización: emparejamiento, estratificación, ponderación de probabilidad inversa y ajuste de covariable por PS. El emparejamiento y la estratificación por PS, son los métodos más populares en la investigación biomédica ya que son fáciles de entender e implementar computacionalmente. También, el ajuste de covariable por PS es un enfoque favorito, ya que es fácil de aplicar. El enfoque, basado en la ponderación de probabilidad inversa por PS, se usa raramente.

Para la implementación de cada uno de estos métodos, uno de los requerimientos es la estimación de los propensity scores PS_i a partir de los datos de observación y luego a proceder a usar el propensity score estimado en lugar de las covariables X_1, X_2, \dots, X_n para estratificación, emparejamiento y ponderación de las observaciones. En las siguientes secciones explicamos formalmente en qué consisten cada una de estas técnicas.

3.2.1. Matching por el propensity score

Rosenbaum y Rubin (1983a, 1983b) desarrollaron por primera vez el método de propensity score matching, también conocida en la literatura como emparejamiento de puntajes de propensión. El uso del propensity score nos permite ajustar una muestra dada de sujetos tratados y no tratados para imitar una asignación aleatoria de individuos a un grupo de tratamiento y no tratamiento. Por lo tanto, al aplicar técnicas de emparejamiento, los investigadores realizan un "experimento aleatorio hipotético" (Rubin, 1986).

El propensity score matching, es uno de los métodos más potentes y ampliamente utilizados por científicos de todo el mundo para analizar los efectos causales, en estudios observacionales. Consiste en usar un diseño emparejado por el propensity score y emparejar a cada sujeto tratado con otro sujeto no tratado usando el propensity score estimado como variable para seleccionar la pareja (emparejamiento 1:1).

Mas formalmente, el propensity score matching, consiste en crear pares mediante la unificación de sujetos tratados y no tratados que comparten el mismo valor del propensity score o con un propensity score más cercano. Expresamos esta idea como:

$$E^{PS} = \{(T_i^{PS_{T_i}}, \bar{T}_j^{PS_{\bar{T}_j}}) \in T^{PS_T} \times \bar{T}^{PS_{\bar{T}}} | PS_{\bar{T}_j} \text{ es igual a } PS_{T_i} \vee PS_{\bar{T}_j} \text{ es el más cercano a } PS_{T_i}\}$$

Notaciones:

E^{PS} : Muestra emparejada mediante puntaje de propensión (PS).

T : Sujetos tratados

\bar{T} : Sujetos no tratados

PS_T : Puntajes de propensión de los sujetos tratados

$PS_{\bar{T}}$: Puntajes de propensión de los sujetos no tratados

T^{PS_T} : Sujetos tratados asociado con sus puntajes de propensión

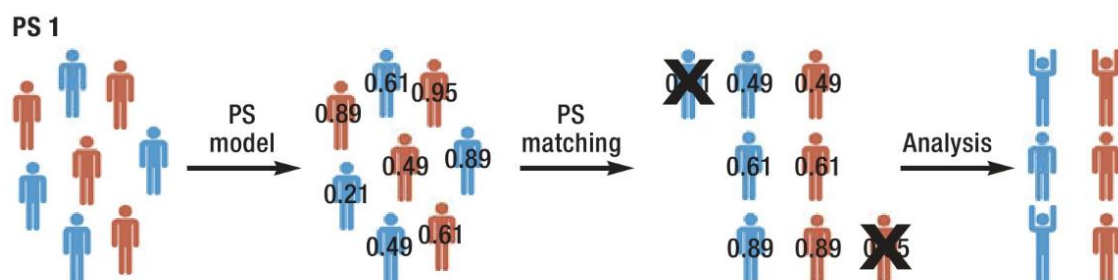
$\bar{T}^{PS_{\bar{T}}}$: Sujetos no tratados asociados con sus puntajes de propensión

PS_{T_i} : Puntaje de propensión para el sujeto tratado i .

$PS_{\bar{T}_j}$: Puntaje de propensión para el sujeto no tratado j .

Para fijar ideas, en la Figura 6 se presenta emparejamiento exacto.

Figura 6: Emparejamiento exacto por el propensity score (Pictograma PS1 en la Figura de Kuss, Blettner & Börgerman)



Sujetos tratados (azul), sujetos no tratados (rojo)

Sujetos no emparejados (etiquetados con una X)

EJEMPLO

Sean:

$$T^{PS_T} = \{T1^{PS_{T1}=0.60}, T2^{PS_{T2}=0.75}, T3^{PS_{T3}=0.81}\},$$

$$\bar{T}^{PS_{\bar{T}}} = \{\bar{T}1^{PS_{\bar{T}1}=0.74}, \bar{T}2^{PS_{\bar{T}2}=0.79}, \bar{T}3^{PS_{\bar{T}3}=0.60}\},$$

Sujetos tratados y no tratados asociados con sus puntajes de propensión respectivamente.

Considerando el orden aleatorio:

$$T1^{PS_{T1}=0.60}, T2^{PS_{T2}=0.75}, T3^{PS_{T3}=0.81}$$

Orden aleatorio de la selección de los sujetos tratados asociados con sus puntajes de propensión, como una idea básica, tenemos la muestra emparejada de los sujetos tratados y no tratados mediante el puntaje de propensión.

$$E^{PS} = \{(T1^{PS_{T1}=0.60}, \bar{T}3^{PS_{\bar{T}3}=0.60}); (T2^{PS_{T2}=0.75}, \bar{T}1^{PS_{\bar{T}1}=0.74}); (T3^{PS_{T3}=0.81}, \bar{T}2^{PS_{\bar{T}2}=0.79})\}$$

- $(T1^{PS_{T1}=0.60}, \bar{T}3^{PS_{\bar{T}3}=0.60})$: El sujeto tratado 1 se ha emparejado con el sujeto no tratado 3, ya que, $PS_{\bar{T}3} = 0.60 = PS_{T1}$
- $(T2^{PS_{T2}=0.75}, \bar{T}1^{PS_{\bar{T}1}=0.74})$: El sujeto tratado 2 se ha emparejado con el sujeto no tratado 1, ya que, $PS_{\bar{T}1} = 0.74$ es el más cercano a $PS_{T2} = 0.75$
- $(T3^{PS_{T3}=0.81}, \bar{T}2^{PS_{\bar{T}2}=0.79})$: El sujeto tratado 3 se ha emparejado con el sujeto no tratado 2, ya que, $PS_{\bar{T}2} = 0.79$ es el más cercano a $PS_{T3} = 0.81$

Usando este método, los sujetos que recibieron el tratamiento son comparados por PS con los sujetos que no recibieron el tratamiento, creando dos grupos que son comparables en posibles factores de confusión, (Rosenbaum y Rubin, 1985).

Sujetos tratados emparejados: $T1^{PS_{T1}=0.60}, T2^{PS_{T2}=0.75}, T3^{PS_{T3}=0.81}$

Sujetos no tratados emparejados: $\bar{T}3^{PS_{\bar{T}3}=0.60}, \bar{T}1^{PS_{\bar{T}1}=0.74}, \bar{T}2^{PS_{\bar{T}2}=0.79}$

3.2.2. Diferentes métodos de matching

El problema con los emparejamientos exactos es que son muy difíciles de obtener, debido a la dimensionalidad de las variables y se vuelven imposibles de controlar si al menos una variable es continua. Rosenbaum y Rubin (1983a, 1983b) propusieron un emparejamiento basado en el puntaje de propensión, en lugar de los factores de confusión, para resolver esta maldición de la dimensionalidad. En la literatura, hay diferentes enfoques de emparejamiento para unir sujetos tratados con sujetos no tratados. Todos ellos intentan crear una muestra emparejada en la que los grupos tratados y no tratados son intercambiables. A continuación, se describe los pasos para realizar el emparejamiento.

PASO 1: Como primer paso es, elegir la relación de sujetos no tratados a sujetos tratados. La relación se refiere a cuántos sujetos no tratados se relacionan con un sujeto tratado, por ejemplo, una relación **1: 1**, significa un sujeto tratado se relaciona con un sujeto no tratado. Mas generalmente, una relación **1: M**, significa un sujeto tratado se relaciona con **M** sujetos no tratados. Por ejemplo, una relación **1: 2**, significa un sujeto tratado se relaciona con dos sujetos no tratados. En este estudio, el emparejamiento se llevará acabo identificando, para cada sujeto tratado un sujeto no tratado, es decir emparejamiento 1:1, por ser el más utilizado.

PASO 2: Como segundo paso es elegir entre emparejamiento sin reemplazo y emparejamiento con reemplazo (Rosenbaum, 2002).

Emparejamiento sin reemplazo: Consiste en seleccionar un sujeto no tratado cualquiera para emparejar con un sujeto tratado, tal que ese sujeto no tratado ya no es candidato para su posible emparejamiento con los sujetos tratados posteriores. Como resultado, permite que cada sujeto no tratado se incluya como máximo en un conjunto emparejado.

Emparejamiento con reemplazo: Consiste en seleccionar un sujeto no tratado cualquiera para emparejar con un sujeto tratado, tal que ese sujeto tratado sigue siendo un candidato para su posible emparejamiento con los sujetos tratados posteriores. Como resultado, permite que el mismo sujeto no tratado pueda estar en múltiples conjuntos emparejados.

Cuando se utiliza el emparejamiento con reemplazo, la estimación de la varianza debe tener en cuenta el hecho de que el mismo sujeto no tratado puede estar en múltiples conjuntos emparejados (Hill y Reiter, 2006). Elegimos emparejamiento (1:1) sin reemplazo, por ser el más utilizado en la investigación médica.

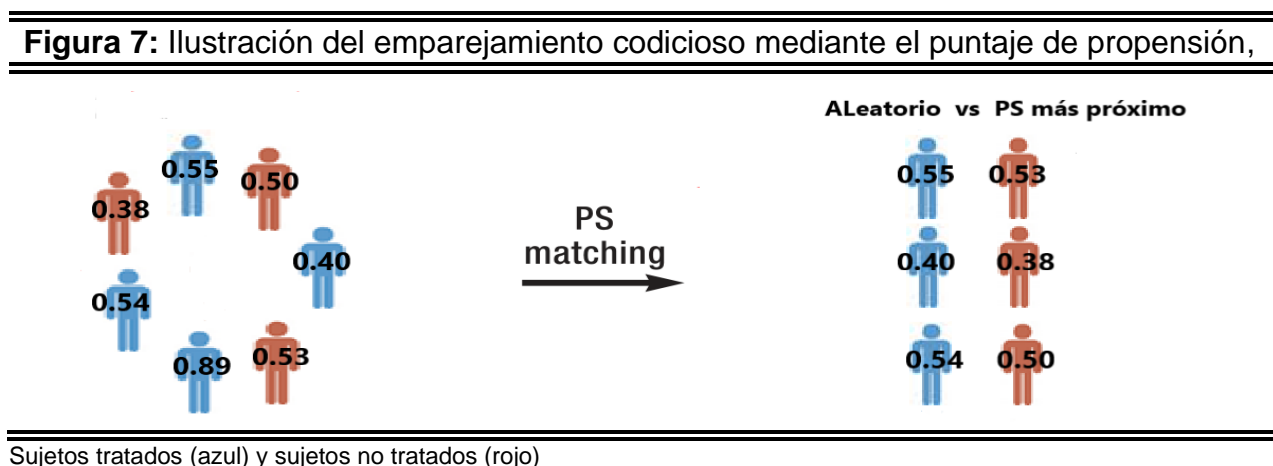
PASO 3: Consiste en la selección de los sujetos tratados aleatoriamente y luego elegir entre el emparejamiento codicioso y óptimo (Rosenbaum, 2002) para la obtención de muestras emparejadas.

Escenario 1: Emparejamiento codicioso

Es un técnico que consiste, en seleccionar aleatoriamente un sujeto del grupo tratado y aparear con un sujeto del grupo no tratado que tiene el propensity score más próximo o cercano. El proceso se repite iterativamente hasta que todos los sujetos del grupo no tratado resulten apareados con todos los sujetos del grupo tratado o hasta que uno quede sujetos tratados para los que se puede encontrar un sujeto no tratado.

$$EC^{PS} = \{(T_i^{PS_{T_i}}, \bar{T}_j^{PS_{\bar{T}_j}}) \in T^{PS_T} \times \bar{T}^{PS_{\bar{T}}} | T_i^{PS_{T_i}} \text{ se extrae al azar} \\ \wedge \bar{T}_j^{PS_{\bar{T}_j}} \text{ se elige} \Leftrightarrow PS_{\bar{T}_j} \text{ es el más cercano a } PS_{T_i}\}$$

En la Figura 7 se presenta un ejemplo del emparejamiento codicioso por el propensity score.



$$EC^{PS} = \{(T1^{PS_{T1}=0.55}, \bar{T}3^{PS_{\bar{T}3}=0.53}); (T2^{PS_{T2}=0.40}, \bar{T}1^{PS_{\bar{T}1}=0.38}); (T3^{PS_{T3}=0.54}, \bar{T}2^{PS_{\bar{T}2}=0.50})\}$$

Es importante notar en la Figura 7, el sujeto no tratado con el propensity score $PS = 0.53$ incluso sería mejor para aparear con el sujeto tratado con $PS = 0.54$, para obtener un emparejamiento más refinado, en el sentido de la producción de muestras más equilibradas.

Dos inconvenientes del emparejamiento codicioso: primero, el orden en que se emparejan los sujetos tratados puede cambiar la calidad de los emparejamientos. Segundo, funciona mal cuando hay mucha competencia por los controles.

Escenario 2: Emparejamiento óptimo

El emparejamiento óptimo se realiza teniendo en cuenta el conjunto total de emparejamientos, es decir, se selecciona emparejamientos individuales del conjunto general de emparejamientos minimizando una medida de distancia global. Según, Austin (2011) el emparejamiento óptimo consiste en minimizar la diferencia absoluta en los puntajes de propensión dentro del conjunto total de emparejamientos. En otras palabras, primero identificamos todos los emparejamientos totales con las sumas totales de distancia de sus puntajes de propensión. El emparejamiento cuya distancia total es el mínimo, será el emparejamiento óptimo.

Si el objetivo es simplemente encontrar grupos bien emparejados, el emparejamiento codicioso resulta suficiente. Sin embargo, si el objetivo son pares bien emparejados, entonces habitualmente se usa el emparejamiento óptimo. Luego de la descripción de los criterios de elección de sujetos tratados, aún no hemos dicho nada sobre los criterios de elección de los sujetos no tratados. En el paso 4 explicamos las técnicas de elección.

PASO 4: Dos técnicas para la selección de sujetos no tratados, primera técnica: selección del vecino más cercano y segunda técnica: selección del vecino más cercano dentro de una distancia de caliper. Elegir entre el emparejamiento del vecino más cercano y emparejamiento dentro de la distancia de caliper especificado, para obtener la muestra emparejada (Rosenbaum y Rubin, 1985).

Escenario 3: Emparejamiento del vecino más cercano

Un sujeto no tratado se aparea con un sujeto tratado que tiene el propensity score (PS) más cercano. Ninguna distancia basada en función del propensity score se considera demasiado grande para realizar el matching. La muestra emparejada se obtiene como:

$$E_{VMC}^{PS} = \{(T_i^{PS_{Ti}}, \bar{T}_j^{PS_{\bar{T}j}}) \in T^{PS_T} \times \bar{T}^{PS_{\bar{T}}} | T_i^{PS_{Ti}} \text{ se extrae al azar,} \\ \bar{T}_j^{PS_{\bar{T}j}} \text{ se elige} \Leftrightarrow PS_{\bar{T}j} \text{ es el más cercano a } PS_{Ti}, \\ d = d(PS_{T2}; PS_{\bar{T}1}) = |PS_{Ti} - PS_{\bar{T}j}| \text{ no se tiene en cuenta} \}$$

E_{VMC}^{PS} : Muestra emparejada mediante el vecino más cercano

T : Sujetos tratados

\bar{T} : Sujetos no tratados

PS_T : Puntaje de propensión de los sujetos tratados

$PS_{\bar{T}}$: Puntaje de propensión de los sujetos no tratados

T^{PS_T} : Sujetos tratados asociados con los puntajes de propensión PS_T .

$\bar{T}^{PS_{\bar{T}}}$: Sujetos no tratados asociados con los puntajes de propensión $PS_{\bar{T}}$.

$T_i^{PS_{T_i}}$: El sujeto tratado i asociado con su puntaje de propensión PS_{T_i} .

$\bar{T}_j^{PS_{\bar{T}_j}}$: El sujeto no tratado j asociado con su puntaje de propensión $PS_{\bar{T}_j}$.

d : Distancia entre puntajes de propensión de los sujetos emparejados.

Para aparear dos individuos mediante el emparejamiento del vecino más cercano, no se imponen restricciones a la diferencia máxima aceptable entre los puntajes de propensión.

EJEMPLO

Sean:

$$T^{PS_T} = \{T_1^{PS_{T_1}=0.80}, T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}\}$$

$$\bar{T}^{PS_{\bar{T}}} = \{\bar{T}_1^{PS_{\bar{T}_1}=0.68}, \bar{T}_2^{PS_{\bar{T}_2}=0.64}, \bar{T}_3^{PS_{\bar{T}_3}=0.40}\}$$

Sujetos tratados y no tratados asociados con sus puntajes de propensión respectivamente. Supongamos, que el orden aleatorio de los sujetos tratados:

$$T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}, T_1^{PS_{T_1}=0.80}$$

La muestra emparejada mediante el vecino más cercano es:

$$E_{VMC}^{PS} = \{(T_2^{PS_{T_2}=0.70}, \bar{T}_1^{PS_{\bar{T}_1}=0.68}); (T_3^{PS_{T_3}=0.65}, \bar{T}_2^{PS_{\bar{T}_2}=0.64}); (T_1^{PS_{T_1}=0.80}, \bar{T}_3^{PS_{\bar{T}_3}=0.40})\}$$

Notemos que,

$$d(PS_{T_2}; PS_{\bar{T}_1}) = |PS_{T_2} - PS_{\bar{T}_1}| = |0.70 - 0.68| = 0.02$$

$$d(PS_{T_3}; PS_{\bar{T}_2}) = |PS_{T_3} - PS_{\bar{T}_2}| = |0.65 - 0.64| = 0.01$$

$$d(PS_{T_1}; PS_{\bar{T}_3}) = |PS_{T_1} - PS_{\bar{T}_3}| = |0.80 - 0.40| = 0.40$$

Es importante destacar que a pesar de que la distancia entre PS_{T_1} y $PS_{\bar{T}_3}$ igual a 0.40 es un número demasiado grande comparado con 0.02 y 0.01, el emparejamiento se realiza sin restricciones de distancia, ya que, según el método del emparejamiento del vecino más cercano ninguna distancia de PS definida se considera demasiado grande para el emparejamiento.

Si varios sujetos no tratados tienen puntajes de propensión que son igualmente cercanos del sujeto tratado, uno de estos sujetos no tratados se selecciona al azar.

EJEMPLO.

Sean:

$$T^{PS_T} = \{T_1^{PS_{T_1}=0.80}, T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}\},$$

$$\bar{T}^{PS_{\bar{T}}} = \{\bar{T}_1^{PS_{\bar{T}_1}=0.68}, \bar{T}_2^{PS_{\bar{T}_2}=0.64}, \bar{T}_3^{PS_{\bar{T}_3}=0.76}, \bar{T}_4^{PS_{\bar{T}_4}=0.68}, \bar{T}_5^{PS_{\bar{T}_5}=0.68}\}$$

Sujetos tratados y no tratados asociados con sus puntajes de propensión. Supongamos el orden aleatorio de los sujetos tratados es dada por:

$$T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}, T_1^{PS_{T_1}=0.80}$$

En el ejemplo existen tres sujetos no tratados $\{\bar{T}_1, \bar{T}_4, \bar{T}_5\}$ cuyos puntajes de propensión son igualmente cercanos del sujeto tratado T_2 . Entonces se extrae aleatoriamente un sujeto no tratado de $\{\bar{T}_1, \bar{T}_4, \bar{T}_5\}$, digamos se extrae \bar{T}_5 . Por lo tanto, la muestra emparejada mediante el emparejamiento del vecino más cercano es:

$$E_{VMC}^{PS} = \{(T_2^{PS_{T_2}=0.70}, \bar{T}_5^{PS_{\bar{T}_5}=0.68}); (T_3^{PS_{T_3}=0.65}, \bar{T}_2^{PS_{\bar{T}_2}=0.64}); (T_1^{PS_{T_1}=0.80}, \bar{T}_3^{PS_{\bar{T}_3}=0.76})\}$$

Escenario 4: Emparejamiento dentro de una distancia de calibre

Esta técnica es la más usada, consiste en el emparejamiento limitado dentro una distancia definido por el investigador. Un sujeto que no tiene una pareja dentro de la distancia de caliper se excluye del análisis. La distancia de caliper óptima recomendada es la quinta parte (0.2) de la desviación estándar de logit del propensity score.

El funcionamiento de esta técnica es similar del emparejamiento del vecino más cercano con una restricción adicional de que la diferencia absoluta entre los puntajes de propensión de los sujetos apareados debe estar dentro de la distancia de caliper, es decir, son todos los sujetos emparejados mediante el emparejamiento del vecino más cercano talque la diferencia absoluta entre los puntajes de propensión es menor que la distancia de caliper. La muestra emparejada mediante este método, se obtiene como:

$$E_{D_{calibre}}^{VMC} = (T_i^{PS_{T_i}}, \bar{T}_j^{PS_{\bar{T}_j}}) \in T^{PS_T} \times \bar{T}^{PS_{\bar{T}}}$$

$T_i^{PS_{T_i}}$: Se extrae aleatoriamente

$\bar{T}_j^{PS_{\bar{T}_j}}$: Se elige, si cumple dos condiciones:

- Pertenece al conjunto restringido (CR) de todos los sujetos no tratados con puntajes de propensión más cercanas del sujeto tratado dado y cuya diferencia absoluta en los puntajes de propensión está por debajo de un umbral pre especificado (la distancia de calibre).

$$CR = \{\bar{T}_j^{PS_{\bar{T}_j}} \in \bar{T}^{PS_{\bar{T}}} | PS_{\bar{T}_j} \text{ es el más cercano a } PS_{T_i} \\ \wedge d(PS_{T_i}, PS_{\bar{T}_j}) = |PS_{T_i} - PS_{\bar{T}_j}| < D_{calibre}\}$$

- Minimiza la diferencia absoluta en los puntajes de propensión dentro del conjunto restringido de sujetos no tratados.

$$|PS_{T_i} - PS_{\bar{T}_j}| \text{ alcanza el minimo valor en } PS_{\bar{T}_j}, \bar{T}_j^{PS_{\bar{T}_j}} \in CR$$

$E_{D_{\text{calibre}}}^{\text{VMC}}$: muestra emparejada mediante el emparejamiento del vecino más cercano dentro de una distancia de caliper especificada

D_{calibre} : es la distancia de calibre especificado por el investigador.

En la práctica, hacemos lo siguiente: Dado un sujeto tratado $T_i^{PS_{T_i}}$, seleccionamos a todos los sujetos no tratados cuyo puntaje de propensión se encuentra dentro de una distancia especificada del sujeto tratado.

$$CR = \{\bar{T}_j^{PS_{\bar{T}_j}} \in \bar{T}^{PS_{\bar{T}}} \mid |PS_{T_i} - PS_{\bar{T}_j}| < D_{\text{calibre}}\}$$

De este conjunto restringido de sujetos no tratados, el sujeto no tratado cuyo puntaje de propensión es el más cercano del sujeto tratado, es seleccionado para aparear con el sujeto tratado.

$$\bar{T}_j^{PS_{\bar{T}_j}} \in CR \text{ tal que } PS_{\bar{T}_j} \approx PS_{T_i} \Rightarrow (T_i^{PS_{T_i}}, \bar{T}_j^{PS_{\bar{T}_j}}) \in E_{D_{\text{calibre}}}^{\text{VMC}}$$

Nota: Se ha introducido la notación $PS_{\bar{T}_j} \approx PS_{T_i}$ para denotar que: $PS_{\bar{T}_j}$ es el más cercano a PS_{T_i} .

Si $CR = \emptyset$, entonces no hay sujetos no tratados en el conjunto restringido para aparear con el sujeto tratado $T_i^{PS_{T_i}}$.

EJEMPLO. Sean:

$$T^{PS_T} = \{T_1^{PS_{T_1}=0.80}, T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}\},$$

$$\bar{T}^{PS_{\bar{T}}} = \{\bar{T}_1^{PS_{\bar{T}_1}=0.68}, \bar{T}_2^{PS_{\bar{T}_2}=0.64}, \bar{T}_3^{PS_{\bar{T}_3}=0.30}\}$$

Sujetos tratados y no tratados con puntajes de propensión respectivamente.

Supongamos el orden aleatorio de selección de los sujetos tratados es:

$$T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}, T_1^{PS_{T_1}=0.80}$$

Para obtener la muestra emparejada mediante esta técnica, primero, se define la distancia de caliper, por ejemplo, ($D_{\text{caliper}} = 0.2$)

Supongamos que los sujetos tratados se han seleccionado mediante el orden aleatorio: $T_2^{PS_{T_2}=0.70}, T_3^{PS_{T_3}=0.65}, T_1^{PS_{T_1}=0.80}$, ahora, identificamos a todos los sujetos no tratados cuyo puntaje de propensión se encuentra dentro de la distancia de caliper.

$$CR = \{\bar{T}_j^{PS_{\bar{T}_j}} \in \bar{T}^{PS_{\bar{T}}} \mid |PS_{T_i} - PS_{\bar{T}_j}| < D_{\text{calibre}} = 0.2\}$$

▪ Para el sujeto tratado $T_2^{PS_{T_2}=0.70}$, Tenemos:

$$|PS_{T_2} - PS_{\bar{T}_j}| = |0.70 - PS_{\bar{T}_j}| < 0.2$$

$$|0.70 - PS_{\bar{T}_1}| = |0.70 - 0.68| = 0.02 < 0.2, \text{ se cumple}$$

$$|0.70 - PS_{\bar{T}_2}| = |0.70 - 0.64| = 0.06 < 0.2, \text{ se cumple}$$

$$|0.70 - PS_{\bar{T}_3}| = |0.70 - 0.30| = 0.40 < 0.2, \text{ no se cumple}$$

$CR = \{\bar{T}1^{PS_{\bar{T}1}=0.68}; \bar{T}2^{PS_{\bar{T}2}=0.64}\}$. Entonces, $PS_{\bar{T}1} = 0.68$ es el más cercano a $PS_{T2} = 0.70$. Entonces se obtiene:

$$(T2^{PS_{T2}=0.70}, \bar{T}1^{PS_{\bar{T}1}=0.68}) \in E_{D_{calibre}}^{VMC}$$

▪ Para el sujeto tratado $T3^{PS_{T3}=0.65}$, Tenemos:

$$|PS_{T3} - PS_{\bar{T}j}| = |0.65 - PS_{\bar{T}j}| < 0.2$$

$$|0.65 - PS_{\bar{T}2}| = |0.65 - 0.64| = 0.01 < 0.2, \text{ se cumple}$$

$$|0.65 - PS_{\bar{T}3}| = |0.65 - 0.30| = 0.35 < 0.2, \text{ no se cumple}$$

$CR = \{\bar{T}2^{PS_{\bar{T}2}=0.64}\}$. Entonces, $\bar{T}2^{PS_{\bar{T}2}=0.64}$ es el único sujeto no tratado con $PS_{\bar{T}2} = 0.64$ el más cercano a $PS_{T3} = 0.65$. Entonces, se obtiene:

$$(T3^{PS_{T3}=0.65}, \bar{T}2^{PS_{\bar{T}2}=0.64}) \in E_{D_{calibre}}^{VMC}$$

▪ Para el sujeto tratado $T1^{PS_{T1}=0.80}$, Tenemos:

$$|PS_{T1} - PS_{\bar{T}j}| = |0.80 - PS_{\bar{T}j}| < 0.2$$

$$|0.80 - PS_{\bar{T}3}| = |0.80 - 0.30| = 0.50 < 0.2, \text{ no se cumple}$$

$CR = \emptyset$. Entonces, el sujeto tratado $T1^{PS_{T1}=0.80}$ no se aparea con ningún sujeto no tratado.

Por lo tanto, la muestra emparejada mediante el método del vecino más cercano dentro de una distancia de caliper es:

$$E_{D_{calibre}}^{VMC} = \{(T2^{PS_{T2}=0.70}, \bar{T}1^{PS_{\bar{T}1}=0.68}); (T3^{PS_{T3}=0.65}, \bar{T}2^{PS_{\bar{T}2}=0.64})\}$$

Escenario 5: Emparejamiento de caliper por logit (PS).

Requiere tener en cuenta algunos tecnicismos: Cuando se usa el emparejamiento de caliper, no existe una definición uniformemente acordada de lo que constituye una distancia máxima aceptable (distancia de caliper) de hecho, en la literatura médica, se ha utilizado una amplia gama de distancias de caliper. En 1973, Cochran y Rubin, examinaron la reducción en el sesgo debido a una sola variable de confusión normalmente distribuida. Realizaron emparejamiento por esta variable de confusión con distancias de caliper que eran proporcionales a la desviación estándar de la variable de confusión, es decir, emplearon como distancia de caliper:

$$D_{calibre} = S_{c_T} / S_{c_{\bar{T}}}$$

Donde, S_{c_T} y $S_{c_{\bar{T}}}$ denotan la desviación estándar de la variable confusora en sujetos tratados y no tratados, respectivamente. El uso de la distancia de caliper igual a $S_{c_T} / S_{c_{\bar{T}}}$ fue uno de los hallazgos más importante en aquella época, ya que, este descubrimiento permitió a los científicos eliminar 99% del sesgo debido a los factores de confusión.

Basado en los hallazgos y resultados obtenidos por Cochran y Rubin, en la literatura, existen argumentos teóricos para realizar emparejamiento dentro de la distancia de caliper proporcional a la desviación estándar de logit del propensity score, ya que es más probable que el logit del propensity score siga una distribución normal.

$$D_{\text{calibre}} = k \times S_{\text{logit}(PS)}$$

Donde, $S_{\text{logit}(PS)}$ denotan la desviación estándar de logit del propensity score. Además,

$$\text{logit}(PS) \sim \mathcal{N}(\mu_{\text{logit}(PS)}; \sigma_{\text{logit}(PS)}^2) \quad (\text{Aproximadamente})$$

$$\text{logit}(PS) = \log \frac{PS}{(1 - PS)}$$

En 1985, Rosenbaum y Rubin, descubrieron que cuando la varianza de logit del propensity score para los sujetos tratados es la misma que la varianza para los sujetos no tratados, el uso de la distancia de caliper igual a 0.2 de la desviación estándar agrupada de logit del propensity score elimina aproximadamente 99% del sesgo debido a los factores de confusión.

Si $\sigma_{\text{logit}(PS_T)}^2 = \sigma_{\text{logit}(PS_{\bar{T}})}^2$. Entonces,

$$D_{\text{calibre}} = \left(\sqrt{\frac{S_{\text{Logit}(PS_T)}^2 + S_{\text{Logit}(PS_{\bar{T}})}^2}{2}} \right) \times 0.2$$

Donde, $S_{\text{Logit}(PS_T)}^2$ y $S_{\text{Logit}(PS_{\bar{T}})}^2$ denotan la varianza de la muestra de logit del propensity score en los sujetos tratados y no tratados, respectivamente.

$$\sqrt{\frac{S_{\text{Logit}(PS_T)}^2 + S_{\text{Logit}(PS_{\bar{T}})}^2}{2}} = \text{Desviación estándar agrupada de } \text{logit}(PS)$$

Por otro lado, Austin (2011b) examinó las distancias de caliper óptimas al estimar las diferencias de riesgo y las diferencias en las medias haciendo sugerencia que los investigadores usen distancia de caliper igual a 0.2.

Escenario 6: Emparejamiento alternativo

Otro enfoque utilizado frecuentemente para obtener muestras emparejadas es hacer coincidir los sujetos tratados con un valor s del propensity score estimado con los sujetos no tratados que tienen un valor $s \pm 0.05$, o alguna otra pequeña diferencia de proximidad (Hernán y Robins, 2020). Por ejemplo, un sujeto tratado con propensity score estimado igual a 0.6563 podría coincidir con un sujeto no tratado cuyo propensity score es igual a 0.6579.

Según, Hernán y Robins, “la definición de cercanía o proximidad en el propensity matching implica una comparación de sesgo y varianza”. Además, afirman que la definición de cercanía también está relacionada con la de positividad.

Si los criterios de cercanía son demasiados imprecisos, los sujetos con valores relativamente diferentes del propensity score se compararán entre sí, la distribución de las covariables no se equilibrará suficientemente entre los grupos tratados y no tratados en la muestra emparejada, y la intercambiabilidad no se mantendrá.

Por otro lado, si los criterios de cercanía son demasiados estrictos y varios sujetos quedan excluidas al realizar el emparejamiento, habría intercambiabilidad aproximada, pero la estimación del efecto del tratamiento puede tener intervalos de confianza del **95%** más amplios.

Una de las ventajas de realizar el emparejamiento por el propensity score es la capacidad de mostrar explícitamente las características registradas de los pacientes tratados y no tratados, que nos permite poder evaluar si la distribución de las características es similar en pacientes tratados y no tratados. Una de las desventajas al realizar el emparejamiento por PS, se excluye a los pacientes para los que no se puede encontrar un compañero de emparejamiento. Esto puede dar como resultado números más bajos y, por lo tanto, menos poder estadístico del emparejamiento por el propensity score.

3.2.2.1. Estudio de la simulación de los diferentes métodos PSM

Debido a que en la literatura se ha realizado muchas simulaciones para los diferentes métodos del emparejamiento ya no ha sido necesario hacer simulaciones de Monte Carlo para mirar el rendimiento de diferentes algoritmos de matching basado en el propensity score. En las literaturas publicadas el método del emparejamiento en el logit del PS, dentro de una distancia de caliper es el más utilizado y recomendado, debido a que el logit del propensity score tiene mayor posibilidad de seguir la ley de la distribución normal. Aunque otras simulaciones optan por la ponderación inversa del propensity score ante efectos no uniformes del tratamiento. Cuando se dispone de muestras grandes, el ajuste por el propensity score es la más utilizada.

3.2.2.2. Verificación de equilibrio en la distribución de covariables

Ya que, los métodos del propensity score se usan con el objetivo de hacer frente a los desequilibrios en las distribuciones de covariables entre los grupos de tratamiento. La verificación de equilibrio se hace para comprobar aproximada la intercambiabilidad y es importante realizarlo, para que las inferencias causales sean posibles. Las verificaciones de equilibrio se realizan antes y después del emparejamiento y estratificación por el propensity score mediante gráficos, pruebas estadísticas y diferencias estandarizadas.

3.2.2.3. Verificación de equilibrio mediante gráficos

Para las covariables categóricas, podemos usar gráficos de barras apiladas y para las variables continuas histogramas. Adicionalmente, se compara la proporción de individuos entre los grupos de tratamiento en caso de las covariables categóricas, y para las covariables continuas se comparan las medias. Las comparaciones se deben realizar antes y después de realizar el emparejamiento o estratificación por el propensity score.

3.2.2.4. Verificación de equilibrio con pruebas estadísticas

Para la verificación de equilibrio mediante pruebas estadísticas se puede aplicar, la igualdad de medias entre los grupos de tratamiento con una prueba t para covariables continuas, como propusieron originalmente Rosenbaum y Rubin (1985). Sin embargo, este enfoque tiene una gran desventaja ya que las pruebas t dependen del tamaño de la muestra. El emparejamiento por el propensity score a menudo reduce el tamaño de la muestra, además, diferentes técnicas de emparejamiento conducen a diferentes tamaños de muestra. Por lo tanto, pueden producirse resultados de prueba no significativas. Las pruebas de equilibrio basadas en la significancia también son controversiales porque el equilibrio es una propiedad de la muestra, no de la población. Para las covariables categóricas se utiliza una prueba χ^2 . Las pruebas se aplican a los datos antes como después del emparejamiento o estratificación.

3.2.2.5. Verificación de equilibrio con diferencias estandarizadas

Calcular la diferencia estandarizada después del emparejamiento y estratificación por el propensity score también es una prueba de equilibrio común (Rosenbaum y Rubin 1985).

Para una covariable continua, la diferencia estandarizada se define como:

$$d = \frac{(\bar{X}_{tratado} - \bar{X}_{no\ tratado})}{\sqrt{\frac{S^2_{tratado} - S^2_{no\ tratado}}{2}}}$$

Donde $\bar{X}_{tratado}$ y $\bar{X}_{no\ tratado}$ son la media de la muestra de la covariable en los sujetos tratados y no tratados respectivamente, mientras que $S^2_{tratado}$ y $S^2_{no\ tratado}$ son las varianzas muestrales de la covariable en los sujetos tratados y no tratados respectivamente.

Para las variables dicotómicas, la diferencia estandarizada se define como:

$$d = \frac{(\hat{P}_{tratado} - \hat{P}_{no\ tratado})}{\sqrt{\frac{\hat{P}_{tratado} (1 - \hat{P}_{tratado}) + \hat{P}_{no\ tratado} (1 - \hat{P}_{no\ tratado})}{2}}}$$

Donde $\hat{P}_{tratado}$ y $\hat{P}_{no\ tratado}$ denotan la prevalencia o la media de la variable dicotómica en sujetos tratados y no tratados, respectivamente. La diferencia estandarizada es la diferencia de medias en los grupos tratado y no tratado en unidades de la desviación estándar agrupada. Además, la diferencia estandarizada no está influenciada por el tamaño de la muestra y permite la comparación de equilibrio relativo de las variables medidas en diferentes unidades.

En su trabajo original, Rosenbaum y Rubin (1985) consideran que el 20% representa un gran sesgo.

Una diferencia estándar menor al 10% indica una diferencia insignificante en la media o prevalencia de una covariable entre los grupos de tratamiento (Normand et al., 2001). Sin embargo, actualmente, se considera que los sesgos por debajo del 3-5% indican un equilibrio suficiente en los análisis del emparejamiento (Caliendo y Kopeinig 2008).

3.2.2.6. Estimación de efectos de tratamiento basado en matching

Para realizar inferencias causales nos basamos en las afirmaciones de Hernán y Robins (2020), “bajo la intercambiabilidad y la positividad dada el propensity score, las medidas de asociación en la población emparejada son estimaciones consistentes de las medidas de efecto, por ejemplo, la razón de riesgo asociativo en la población emparejada estima consistentemente la razón de riesgo causal en la población emparejada”. A partir de la muestra emparejada se puede calcular dos medidas de los efectos del tratamiento:

1) El efecto promedio del tratamiento (ATE)

$$ATE = E[Y^{A=1}] - E[Y^{A=0}] = \frac{\sum_{i=n}^n Y_i^{A=1}}{n} - \frac{\sum_{i=n}^n Y_i^{A=0}}{n}$$

(El ATE es el efecto promedio, a nivel de la población)

2) El efecto promedio del tratamiento para el tratado (ATT)

$$ATT = E[Y^{A=1}] - E[Y^{A=0}]$$

(El ATT es el efecto promedio del tratamiento en aquellos sujetos que finalmente recibieron el tratamiento)

En la práctica el matching por el propensity score no produce una muestra perfectamente emparejado. Rosenbaum y Rubin (1985) recomendaron ajustar las covariables no balanceadas, denotadas como (X_1^*, \dots, X_q^*) en un modelo de regresión para eliminar sesgo de selección restante después de realizar el matching.

$$E[Y|A, X_1^*, \dots, X_q^*] = \alpha + \beta_0 A + \beta_1 X_1^* + \dots + \beta_q X_q^* \quad (1.1)$$

Del modelo ajustado (1.1) se obtienen: $\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_q$

$$Y^{A=a} = \hat{\alpha} + \hat{\beta}_0 A + \hat{\beta}_1 X_1^* + \dots + \hat{\beta}_q X_q^* \quad a \in \{0, 1\}$$

$$Y^{A=1} = \hat{\alpha} + \hat{\beta}_0 + \hat{\beta}_1 X_1^* + \dots + \hat{\beta}_q X_q^*$$

$$Y^{A=0} = \hat{\alpha} + \hat{\beta}_1 X_1^* + \dots + \hat{\beta}_q X_q^*$$

Ahora el efecto promedio del tratamiento ATT sería calculado como:

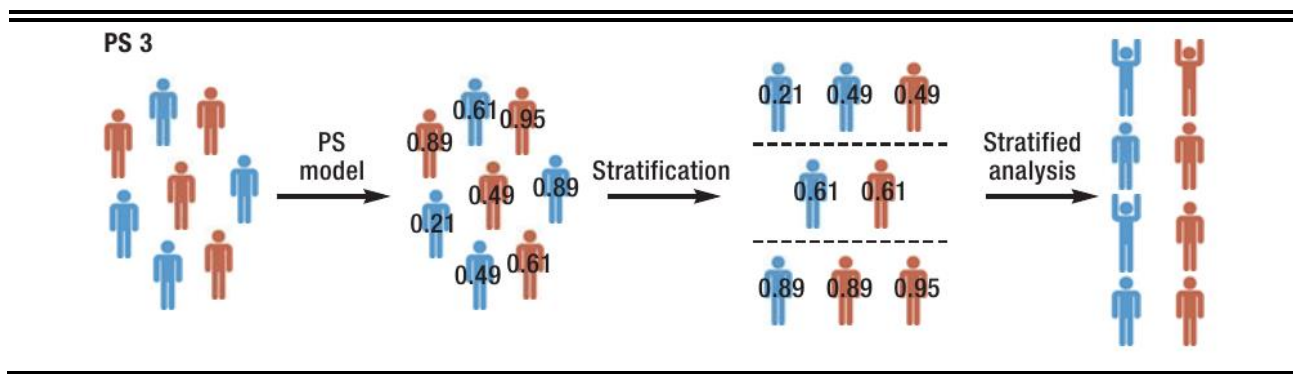
$$ATT = E[Y^{A=1}] - E[Y^{A=0}] = \hat{\beta}_0$$

Por lo tanto, para el emparejamiento del vecino más cercano, el emparejamiento dentro de la distancia de caliper, el emparejamiento métrico de Mahalanobis y similares que producen emparejamientos uno a uno, el ATT se puede obtener ajustando las covariables no balanceados después del matching en el modelo de regresión (1.1).

3.2.3. Estratificación por el propensity score

El uso del propensity score para la estratificación de las observaciones implica clasificar las observaciones en función del propensity score estimado y luego estratificar en subconjuntos o estratos basados en umbrales previamente definidos a partir del propensity score. Dentro de cada estrato, algunos sujetos habrán recibido el tratamiento de interés otros no, pero todos tendrán unas probabilidades estimadas similares de recibirlo dadas sus covariables observadas. En la Figura 8 se presenta un ejemplo de estratificación en tertiles del propensity score.

Figura 8: Estratificación en tertiles del propensity score (Pictograma PS 3 en la Figura de Kuss, Blettner & Börgerman)



Sujetos tratados (azul) y sujetos no tratados (rojo).
Estratificados en tertiles.

La estratificación agrupa a individuos que tienen valores similares o iguales para la variable de estratificación. El propensity score es la variable de estratificación y su distribución se usa para definir los estratos. Dos escenarios de estratificación haciendo uso del propensity score: quintiles y deciles del propensity score.

ESCENARIO 1: Un enfoque ampliamente utilizado es dividir las observaciones en cinco grupos de igual tamaño usando los quintiles del propensity score estimado. Cuando se usa el propensity score como una variable de estratificación, el propensity score debe dividirse como mínimo en cinco estratos. Dentro de cada estrato, las distribuciones de las covariables medidas estarán suficientemente equilibradas entre los grupos de tratamiento. El uso de los quintiles tiene sustento científico, gracias a los estudios y demostraciones de dos mentes brillantes.

- Cochran (1968) demostró que la estratificación en los quintiles de una variable confusora y continua elimino aproximadamente el 99% del sesgo debido a esa variable.
- Rosenbaum y Robin (1984) extendieron este resultado a la estratificación por propensity score, mostrando que estratificar en los quintiles del propensity score estimado elimina aproximadamente el 90% del sesgo debido a factores de confusión medidos al estimar el efecto de un tratamiento lineal.

Estimación de efectos de tratamiento basado en la estratificación

A partir de la muestra estratificada, el efecto del tratamiento A sobre el resultado Y puede estimarse comparando los resultados directamente entre sujetos tratados y no tratados dentro de cada estrato. Las estimaciones específicas del efecto del tratamiento en los estratos se agrupan para estimar un efecto general del tratamiento (Rosenbaum y Rubin, 1984). La muestra estratificada disponible, es procesado para estimar la asociación entre la variable de tratamiento y la variable de respuesta dentro de los niveles o estratos aproximadamente balanceados por el propensity score, en teoría, y bajo el supuesto de que no hay más factores de confusión no medidos, se puede alcanzar un escenario que se ha denominado pseudo aleatorización, en el que subgrupos de pacientes comparables constituyen grupos tratados y no tratados que se pueden analizar sin confusión.

Dos formas de obtener el efecto del tratamiento para muestras estratificadas: primero, ponderando la diferencia de medias y segundo, ponderando el efecto del tratamiento en los estratos, obtenidos después del ajuste de covariables no balanceadas en una regresión.

Método 1: Ponderación de la diferencia de medias

Con datos estratificados, el efecto promedio del tratamiento (ATE) es el promedio ponderado de la diferencia de medias por el número total de sujetos tratados y no tratados en los estratos.

$$\widehat{ATE} = \sum_{k=1}^5 \frac{n_{s_k}^{a=1,a=0} (\overline{Y_{s_k}^{a=1}} - \overline{Y_{s_k}^{a=0}})}{N}$$

\widehat{ATE} = Efecto promedio del tratamiento estimado

$\overline{Y_{s_k}^{a=1}}$ = Media de las respuestas para los sujetos tratados

$\overline{Y_{s_k}^{a=0}}$ = Media de las respuestas para los sujetos no tratados

$n_{s_k}^{a=1,a=0}$ = Tamaño de la muestra en el estrato k

N = Tamaño de la muestra.

5 = Número de estratos (estratificación por quintiles)

Para datos estratificados el (ATT): es el promedio ponderado de la diferencia de medias multiplicado por el número de sujetos tratados en cada estrato.

$$\widehat{ATT} = \sum_{k=1}^5 \frac{n_{s_k}^{a=1} (\overline{Y_{s_k}^{a=1}} - \overline{Y_{s_k}^{a=0}})}{N}$$

\widehat{ATT} = Efecto promedio del tratamiento en el tratado (estimado)

$\overline{Y_{s_k}^{a=1}}$ = Media de las respuestas para los sujetos tratados

$\overline{Y_{s_k}^{a=0}}$ = Media de las respuestas para los sujetos no tratados

$n_{s_k}^{a=1}$ = Número total de sujetos tratados en el estrato k (s_k)

N = Tamaño de la muestra.

Método 2: Ponderación mediante regresión

Otra forma de tratar con datos estratificación, consiste en el ajuste de las covariables desequilibradas en la regresión (ecuación 1.1) para cada estrato.

$$E[Y|A, X_1^*, \dots, X_q^*] = \beta_0 + \beta_1 A + \beta_2 X_1^* + \dots + \beta_{q+1} X_q^* \quad (1.1)$$

De la regresión ajustada, se obtiene el efecto promedio del tratamiento estimado para cada estrato k :

$$\widehat{ATE}_{s_k} = \hat{\beta}_{s_k}$$

Para la regresión (1.1) $\hat{\beta}_{s_k}$ es el coeficiente estimado de la variable indicadora del tratamiento A , para cada estrato s_k .

Luego, para estimar el (ATE) y (ATT) calculamos un promedio ponderado de $\hat{\beta}_{s_k}$ multiplicado por el número total de sujetos tratados y no tratados en los estratos (caso ATE), y multiplicado por el número de sujetos tratados en los estratos (caso ATT)

$$\widehat{ATE} = \sum_{k=1}^5 \frac{n_{s_k}^{a=1, a=0} \hat{\beta}_{s_k}}{N}$$
$$\widehat{ATT} = \sum_{k=1}^5 \frac{n_{s_k}^{a=1} \hat{\beta}_{s_k}}{N}$$

ESCENARIO 2: Otra opción con respecto a la elección de los quintiles son los deciles del propensity score estimado. Según los deciles, los sujetos de la población se clasifican en 10 estratos de aproximadamente el mismo tamaño. A partir de la muestra estratificada por deciles, el efecto del tratamiento se estima en cada uno de los 10 estratos. La estimación se realiza de manera análogo caso escenario 1, teniendo en cuenta que la estratificación se hizo usando los deciles del propensity score.

Ponderación de la diferencia de medias:

Efecto promedio del tratamiento ATE :

$$\widehat{ATE} = \sum_{k=1}^{10} \frac{n_{s_k}^{a=1, a=0} (\overline{Y_{s_k}^{a=1}} - \overline{Y_{s_k}^{a=0}})}{N}$$

Efecto promedio del tratamiento en el tratado ATT :

$$\widehat{ATT} = \sum_{k=1}^{10} \frac{n_{s_k}^{a=1} (\overline{Y_{s_k}^{a=1}} - \overline{Y_{s_k}^{a=0}})}{N}$$

Ponderación mediante regresión:

$$\widehat{ATE} = \sum_{k=1}^{10} \frac{n_{s_k}^{a=1, a=0} \hat{\beta}_{s_k}}{N}$$

$$\widehat{ATT} = \sum_{k=1}^{10} \frac{n_{s_k}^{a=1} \hat{\beta}_{s_k}}{N}$$

Para estimar (*ATE*) y (*ATT*), se recomienda usar ponderación de la diferencia de medias, cuando el tamaño de la muestra en cada estrato es relativamente pequeño. Para muestras relativamente grandes es preferible usar ponderación de los coeficientes de regresión $\hat{\beta}_{s_k}$.

3.2.4. Ponderación por el propensity score

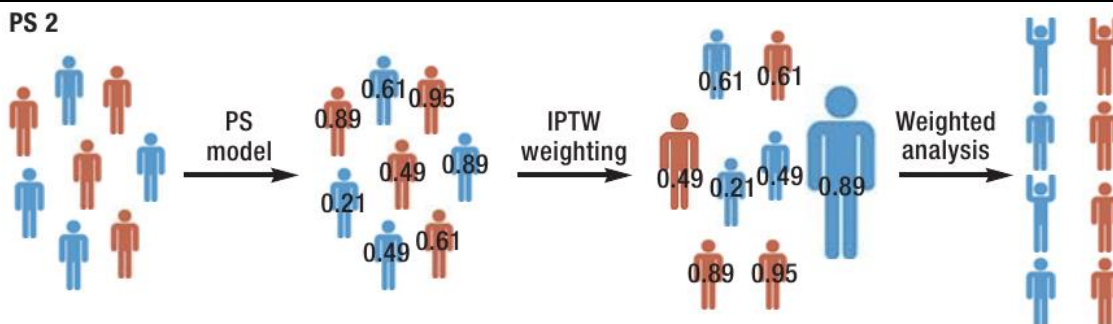
Ponderación por el propensity score consiste en la asignación de pesos a los individuos del grupo tratado y no tratado, el recíproco de la probabilidad de tratamiento. Los individuos del grupo tratado reciben el peso $1 / PS$, y los individuos del grupo no tratado reciben el peso $1 / (1 - PS)$. Se puede interpretar de manera intuitiva de que un paciente tratado con el propensity score bajo recibe una alta ponderación porque es similar a un paciente no tratado con el propensity score bajo, por lo que se puede hacer una comparación válida entre los dos pacientes.

Los pesos para cada individuo se definen como:

$$W_i = \frac{A_i}{PS_i} + \frac{1 - A_i}{1 - PS_i}$$

El denominador es PS_i para los individuos tratados ($A = 1$) y $1 - PS_i$ para los individuos no tratados ($A = 0$). Un inconveniente de la asignación de los pesos es que se podría dar un peso excesivo a algunas observaciones que podrían influir dramáticamente en los resultados. En la Figura 9 se presenta un ejemplo de ponderación por el propensity score.

Figura 9: Ponderación por el propensity score (Pictograma PS 2 en la Figura de Kuss, Blettner & Börgerman)



Sujetos tratados (rojo) y sujetos no tratados (azul).

Los sujetos con peso más alto son más grandes en la figura.

Una de las fortalezas más importantes de la ponderación por el propensity score es la obtención de la muestra ponderada en la que la distribución de las covariables medidas es independiente de la asignación al tratamiento.

Estimación de efectos de tratamiento basado en ponderación

A partir de la muestra ponderada por la inversa del propensity score, el efecto del tratamiento se obtiene bajo las condiciones de identificabilidad, es decir, bajo los siguientes supuestos: que todos los factores de confusión hayan sido correctamente medidos e introducidos en el modelo del propensity score, concretamente bajo las condiciones de identificabilidad. La muestra ponderada, se puede usar de dos maneras para estimar el efecto del tratamiento. Se puede estimar directamente ATE en toda la información original como

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N \left[\frac{A_i Y_i}{PS_i} - \frac{(1 - A_i) Y_i}{1 - PS_i} \right]$$

Es el estimador de Horvitz y Thompson (1952). N_1 y N_0 es el número total de individuos tratados y no tratados, respectivamente, en los datos originales completos ($N_1 + N_0 = N$).

También la muestra ponderada se puede ajustar las covariables desequilibradas en una regresión (Ecuación 1.1), con pesos de regresión

$$W_i = \frac{A_i}{PS_i} - \frac{(1 - A_i)}{1 - PS_i}$$

El (*ATE*) *se estima* como $\widehat{ATE} = \hat{\beta}_1$ (Robins et al., 2000); o con pesos como

$$W_i = A_i + \frac{(1 - A_i)PS_i}{1 - PS_i}$$

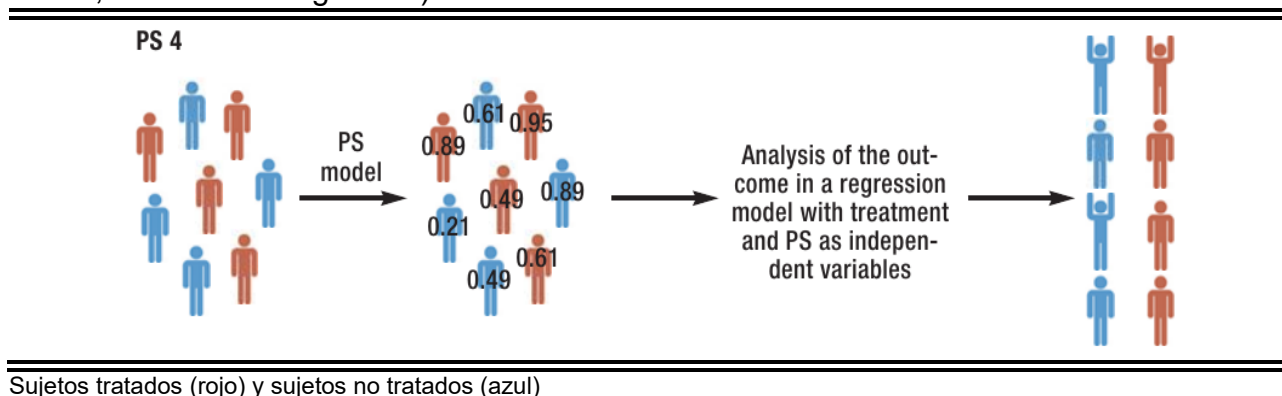
El (*ATT*) se estima como $\widehat{ATT} = \hat{\beta}_1$ (Hirano, Imbens y Knight, 2003; Morgan y Todd, 2008).

Al realizar las estimaciones con la muestra ponderada, los intervalos de confianza a los 95% calculados en los modelos del desenlace ponderados por la inversa del propensity score pueden dar lugar a intervalos que no cubren correctamente el 95%. En la literatura, sugieren usar en su lugar, estimadores con varianza robusta o Bootstrap no paramétrico.

3.2.5. Propensity score como variable de ajuste

El propensity score como un número único que integra todos los factores (X_1, \dots, X_n) asociados al tratamiento A , nos permite tomar los n factores (X_1, \dots, X_n) y condensarlas en un solo factor PS . Usamos ese único factor PS evitando ajustar por los n factores (X_1, \dots, X_n) . Esta forma peculiar de ajuste, resulta muy eficiente para hacer inferencias causales en los estudios observacionales. Por ejemplo, bajo las condiciones de que el propensity score es un predictor lineal, la estrategia de ajuste se introduce en un modelo convencional de regresión lineal, modelos ANCOVA, logística, de Poisson o de Cox. En la Figura 10 se presenta un ejemplo gráfico para fijar ideas.

Figura 10: Propensity score como variable de ajuste (Pictograma PS 4 en la Figura de Kuss, Blettner & Börgerman)



Estimación de efectos de tratamiento ajustado por el propensity score

Bajo el supuesto de que el propensity score, es un predictor lineal, la elección del modelo de regresión depende de la naturaleza de la variable resultado Y .

Si la variable Y es continua, modelizamos con un modelo de regresión lineal.

$$E[Y|X_1, \dots, X_n] = \alpha + \beta X$$

Si la variable resultada Y es dicotómico, modelizamos con un modelo de regresión logística.

$$Pr[Y = 1|X_1, \dots, X_n] = \frac{\exp\{\alpha + \beta X\}}{1 + \exp\{\alpha + \beta X\}}$$

De los cuatro métodos del propensity score, este es el único que requiere que especifique un modelo de regresión que relacione el resultado Y , con los niveles del tratamiento A y el propensity score PS . El ajuste de modelos por el propensity score, es similar a los ajustes tradicionales. Las inferencias sobre el efecto del tratamiento A sobre el resultado Y se derivan de los coeficientes estimados del modelo de regresión ajustado.

- **Para un modelo de regresión lineal:** El efecto promedio del tratamiento A sobre el resultado Y es la diferencia de medias.

En el modelo de regresión lineal se ajustan dos variables: Variable indicadora del tratamiento A y la variable propensity score PS . Obteniendo, el modelo de regresión lineal ajustado.

$$E[Y|A, PS] = \alpha + \beta A + \gamma PS$$

Luego se procede a estimar los coeficientes del modelo lineal ajustado: $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$. Obteniendo, resultados estimados bajo tratamiento ($A = 1$) y no tratamiento $A = 0$.

$$Y^{a=1} = E[Y|A = 1, PS] = \hat{\alpha} + \hat{\beta} + \hat{\gamma}PS$$

$$Y^{a=0} = E[Y|A = 0, PS] = \hat{\alpha} + \hat{\gamma}PS$$

Donde, $Y^{a=1}$ y $Y^{a=0}$, son variables resultado bajo tratamiento y no tratamiento, respectivamente. Entonces, el efecto promedio del tratamiento A , se calcula como la diferencia de medias,

$$E[Y^{a=1}] - E[Y^{a=0}] = \frac{\sum_{i=1}^n \hat{\alpha} + \hat{\beta} + \hat{\gamma}PS_i}{n} - \frac{\sum_{i=1}^n \hat{\alpha} + \hat{\gamma}PS_i}{n} = \hat{\beta}$$

- **Para un modelo de regresión logística:** El efecto del tratamiento A sobre el resultado Y es el *odds ratio* ajustado.

Como primer paso, en el modelo de regresión logística se ajusta dos variables: variable indicadora del tratamiento A y la variable del propensity score (PS). Se obtiene el modelo de regresión logística ajustada.

$$Pr[Y = 1|A, PS] = \frac{\exp\{\alpha + \beta A + \gamma PS\}}{1 + \exp\{\alpha + \beta A + \gamma PS\}}$$

Luego, estimamos los coeficientes del modelo de regresión logística ajustada: $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$. Obteniendo el riesgo para $Y = 1$, bajo tratamiento y no tratamiento:

$$Pr(Y = 1|A = 1, PS) = \frac{\exp\{\hat{\alpha} + \hat{\beta} + \hat{\gamma}PS\}}{1 + \exp\{\hat{\alpha} + \hat{\beta} + \hat{\gamma}PS\}}$$

$$Pr(Y = 1|A = 0, PS) = \frac{\exp\{\hat{\alpha} + \hat{\gamma}PS\}}{1 + \exp\{\hat{\alpha} + \hat{\gamma}PS\}}$$

El efecto promedio del tratamiento A , es el *odds ratio*, y se obtiene como:

$$\widehat{OR}_A = \frac{odds(Y = 1|A = 1, PS)}{odds(Y = 1|A = 0, PS)} = \frac{\exp\{\hat{\alpha} + \hat{\beta} + \hat{\gamma}PS\}}{\exp\{\hat{\alpha} + \hat{\gamma}PS\}} = \exp\{\hat{\beta}\}$$

4. APLICACIÓN DEL PROPENSITY SCORE PARA LA EVALUACIÓN DEL RENDIMIENTO NEUROPSICOLÓGICO EN USUARIOS DE CANNABIS

En este capítulo se aplicarán los métodos del PS a los datos de un estudio real sobre el efecto del consumo regular del cannabis. El cannabis es la droga ilegal más consumida entre los jóvenes de entre 14 y 25 años, con tasas de prevalencia y demandas de tratamiento en aumento constante. En la actualidad, los efectos potencialmente adversos del consumo de cannabis en las funciones ejecutivas siguen siendo controversiales, fomentando la necesidad de nuevos biomarcadores capaces de revelar los déficits neurocognitivos. De acuerdo con diferentes modelos conceptuales, el sistema ejecutivo abarca una gama de funciones disociables, que incluyen atención sostenida, memoria de trabajo, monitoreo de respuesta y toma de decisiones (Robertson y Garavan, 2000; Robbins, 2007; Verdejo-García y Pérez-García, 2007).

4.1. EL ESTUDIO DE VERDEJO ET AL. (2013)

En 2013, un equipo de científicos publicó un artículo sobre un estudio de: “*COMT val158met and 5-HTTLPR Genetic Polymorphisms Moderate Executive Control in Cannabis User*”, que trata sobre los efectos adversos del consumo de cannabis en las funciones ejecutivas, además, investigan dos polimorfismos genéticos que podrían estar moderando los efectos nocivos del consumo de cannabis en la función ejecutiva. El estudio lo realizan suponiendo que los usuarios de cannabis tienen un peor desempeño. Evalúan específicamente cinco diferentes aspectos neurocognitivos mediante la evaluación neuropsicológica computarizada. Para estudiar la relación de los cinco aspectos del rendimiento cognitivo y el consumo de cannabis: el grupo de estudio, el género, la edad, el estado de fumar y el coeficiente intelectual verbal se incluyeron en los modelos ANCOVA. En los resultados obtenidos no encontraron diferencias significativas entre los usuarios de cannabis y los controles en ninguno de los principales índices de rendimiento de las tareas.

Después de la descripción de algunos aspectos más resaltantes del estudio realizado por Verdejo en 2013, a continuación, vamos a realizar el mismo estudio, es decir, la evaluación del rendimiento neuropsicológico en usuarios de cannabis, pero esta vez, utilizando las metodologías del propensity score. En primer lugar, se presenta los objetivos del estudio.

4.2.2. Objetivos del estudio

- i) Evaluar las funciones ejecutivas mediante cuatro métodos diferentes del propensity score en datos que provienen de un estudio observacional.
 - Matching por el propensity score
 - Estratificación por el propensity score
 - Ponderación por el propensity score
 - Ajustar por el propensity score
- ii) La comparación de resultados obtenidos con los métodos del PS, por un lado, y del modelo ANCOVA, por otro lado.

4.2.3. Materiales y métodos

Participantes: El estudio se llevó a cabo con 144 participantes europeos-caucásicos en total, que contiene dos grupos de estudio, 86 participantes de cannabis (grupo cannabis) y 58 participantes no consumidores de drogas (grupo control). En la Tabla 7 se presenta las características demográficas de ambos grupos de estudio.

Tabla 7 Datos descriptivos sobre variables demográficas y características de consumo de droga (Tabla 1 en Verdejo et al., 2013)

	Cannabis (n=86) Media (SD)	Control (n=58) Media (SD)	SMD^a
Edad	21.5 (2.9)	23.2 (4.8)	0.44
Vocabulario WAIS-II	11.3 (2.0)	12.6 (1.9)	0.67
Género			0.08
Masculino	61 (70.9)	39 (67.2)	
Femenino	25 (29.1)	19 (32.8)	
Título universitario, n (%)			0.55
Si	55 (64.0)	50 (87.7)	
No	31 (36.0)	7 (12.3)	
Situación laboral, n (%)			0.15
Empleado	24 (28.2)	22 (38.6)	
Desempleado	23 (27.1)	11 (19.3)	
Estudiante	38 (44.7)	24 (42.1)	
Fumador, n (%)			1.15
Actual fumador	57 (66.3)	9 (16.1)	
No-fumador / ex -fumador	29 (33.3)	47 (83.9)	
Edad al primer consumo de tabaco, años	18.9 (3.0)	19.0 (4.2)	0.03
Cigarrillos por día	9.4 (6.2)	6.6 (5.6)	0.46
Edad al primer consumo de alcohol, años	14.8 (1.4)	15.6 (1.4)	0.59
Años de consumo de alcohol.	6.7 (3.0)	7.6 (4.4)	0.21
Edad al primer consumo de cannabis, años	15.3 (1.5)		
Años de consumo de cannabis	6.2 (2.8)		
Consumo de cannabis de por vida (número de articulaciones)	5645.1 (3908.6)		
Consumo de cannabis durante el año pasado.	1002.2 (678.6)		

^a Valor absoluto de la diferencia de medias estandarizada.

Criterios de selección y exclusión:

Los criterios de inclusión para los consumidores de cannabis fueron: consumo diario de cannabis (>7 porros por semana) durante una duración mínima de 3 años, cumpliendo los criterios del Manual de Diagnóstico y Estadística IV (DSM-IV) y tener al menos educación de nivel básico (>8 años de educación). El criterio de exclusión para los consumidores de cannabis fue basado en el historial toxicológico. Para los participantes sin uso de drogas, los criterios de exclusión adicionales, basados en el historial toxicológico, fueron el uso de cualquier droga hace un año o que hayan consumido más de 10 ocasiones durante la vida.

4.2.4. Procedimiento de la prueba

En el centro de investigación (IMIM-Instituto de Investigación Médica del Hospital del Mar), los participantes fueron sometidos a una evolución médica inicial que incluía: el historial médico detallado, análisis bioquímico, pruebas de drogas en orina y cabello, examen físico y examen neurológico. Posteriormente, los participantes que pasaron la evaluación médica inicial se sometieron a una entrevista psiquiátrica. Todos los protocolos de estudio fueron aprobados y realizados de acuerdo con el comité de ética local (CEIC — Parc de Salut Mar).

4.2.5. Evaluación neuropsicológica

Todos los participantes que cumplieron los criterios de inclusión descritos antes, fueron sometidos a una sesión de evaluación neuropsicológica computarizada de 120 minutos. Se evaluaron cinco aspectos de las funciones ejecutivas, utilizando sondas bien validadas de CANTAB (Robbins et al, 1994, 1998, ver www.camcog.com). Los cinco aspectos evaluados con las medidas de rendimiento obtenidos de las diferentes sondas fueron:

1. **Atención sostenida** (CANTAB Rapid Visual Information Processing Test, RVIP)
Probability of hit (probabilidad de acierto)

La probabilidad de acierto indexa la precisión de atención sostenida y se calculó como: $\text{número de aciertos} / (\text{número de aciertos} + \text{número de errores})$

2. **Memoria de trabajo** (N-back; Watter et al, 2001)
Probability of hit, N=2 (probabilidad de acierto, N=2)

La probabilidad de acierto, N=2, indexa la precisión de la memoria de trabajo y se calculó como: $\text{proporción de aciertos} - \text{proporción de errores}$

3. **Monitoreo de respuesta / cambio de desplazamiento** (CANTAB Intradimensional/Extradimensional set shifting, ID/ED)
Total errors adjusted (total de errores ajustado)

Total, de errores ajustado, es el número total de errores ID / ED, que indexa errores de supervisión del rendimiento.

4. **Planificación/impulsividad-reflexión** (CANTAB Stockings of Cambridge, SOC)
Mean latency to first choice (latencia media a la primera opción)

Latencia media a la primera opción es el tiempo medio de iniciación del SOC, indexa las habilidades de planificación.

5. **Toma de decisiones** (Iowa Gambling Task, IGT; Bechara et al, 2000).
Net score (puntaje neto)

El puntaje neto que indexa la calidad de la toma de decisiones, se calculó como:
suma de opciones de mazo ventajosas (C + D) –
suma de opciones de mazo desventajosas (A + B)

4.2.6. Análisis estadístico

Muestra: Por un lado, en la muestra, las variables numéricas se describen mediante la media y desviación estándar. Por otro lado, las variables categóricas se describen mediante las frecuencias absolutas y relativas. Las diferencias entre ambos grupos de estudio (cannabis vs controles) se describen mediante diferencia de medias estandarizadas.

Modelo estadístico: Para la estimación del propensity score se utilizó la regresión logística multivariable (MLR), que modela el consumo de cannabis en función de las covariables medidas. Los propensity scores estimados son las probabilidades pronosticadas resultantes del consumo de cannabis para cada individuo. Para estudiar la relación entre el rendimiento cognitivo y el consumo de cannabis, los modelos ANCOVA final se ajustaron incluyendo el grupo de estudio como variable predictiva de interés, así como el género, edad, estado de fumar y el coeficiente intelectual verbal estimado IQ, como covariables. Previo al estudio de la relación entre el rendimiento cognitivo y el consumo de cannabis, se realizó el emparejamiento (1:1) por el propensity score utilizando la técnica del vecino más cercano dentro de una distancia de caliper, estratificación en los quintiles del propensity score y la ponderación por el propensity score. La significancia estadística se estableció en 0.05.

Software estadístico: Todos los análisis estadísticos se realizaron en el lenguaje R (2020 The R Foundation for Statistical Computing), versión 4.0.0. Actualmente en el entorno R el paquete no aleatorio nonrandom ya no se encuentra disponible, pero, algunas de las funciones del paquete siguen disponibles en la siguiente dirección:
<https://github.com/cran/nonrandom/tree/master/R#start-of-content>.

Adicionalmente se ha utilizado el paquete MatchIt, actualmente disponible en R, para la obtención de algunos resultados que presentaremos en la aplicación del propensity score en estudio con datos reales.

4.2. APLICACIÓN DEL PROPENSITY SCORE

En esta sección se presenta la implementación en R del propensity score para analizar los datos de un estudio real sobre el efecto del consumo regular del cannabis. Antes de la implementación en R, primero se realiza algunas codificaciones de las variables en los datos del estudio.

Variable predictiva. El grupo de estudio (*group*) como variable predictiva de interés, fue dicotomizado, de la siguiente manera.

(Variable predictiva de interés) $\text{group} = \begin{cases} 0: \text{Cannabis} \\ 1: \text{Control} \end{cases}$

Variable de resultado. La variable de resultado se define en cinco distintas variables neuropsicológicas: Probabilidad of hit (*rvptprob*), Probabilidad of hit N=1 (*nbackdif*), Total errors adjusted (*iedterradj*), Mean latency to firts choise (*otsmlatfirts*), Net score (*igtttotal*) y Z composite score (*zscore*).

$Y = (\text{"rvptprob"}, \text{"nback2dif"}, \text{"iedterradj"}, \text{"otsmlatfirst"}, \text{"igtttotal"}, \text{"zscore"})$

Covariables. Las cuatro covariables posibles son codificados como: género (*gender*), edad (*age*), vocabulario WAIS-III (*civoc*) y el estado de fumar (*smoke*).

(Vector de covariables) $X = (\text{gender}, \text{age}, \text{civoc}, \text{smoke})$

Las covariables *gender* y *smoke* fueron dicotomizados de la siguiente manera:

$\text{gender} = \begin{cases} 0: \text{Male} \\ 1: \text{Female} \end{cases}, \quad \text{Smoke} = \begin{cases} 0: \text{Current smoker} \\ 1: \text{Non smoker/Ex - smoker} \end{cases}$

4.2.1. Missing data

Del conjunto total de datos se eliminaron dos datos missing correspondiente de la covariable “*smoke*” de esta manera nos quedamos con 142 observaciones para estimar el propensity score. Después de la estimación del propensity score, se eliminó un dato missing de la variable “*igtttotal*” de esta manera nos quedamos con 141 observaciones para realizar el emparejamiento, estratificación, ponderación y ajuste mediante el propensity score.

4.2.2. Estimación del propensity score

Para la estimación del propensity score, se ajustan las covariables: **gender**, **age**, **civoc**, **smoke** en el modelo de regresión logística (uso interno de `glm()`) considerando el grupo de estudio (**group**) como variable predictiva de interés. A continuación, se presenta el modelo ajustado para poder ver cómo están relacionadas las covariables con el consumo de cannabis.

$$group \sim gender + age + civoc + smoke$$

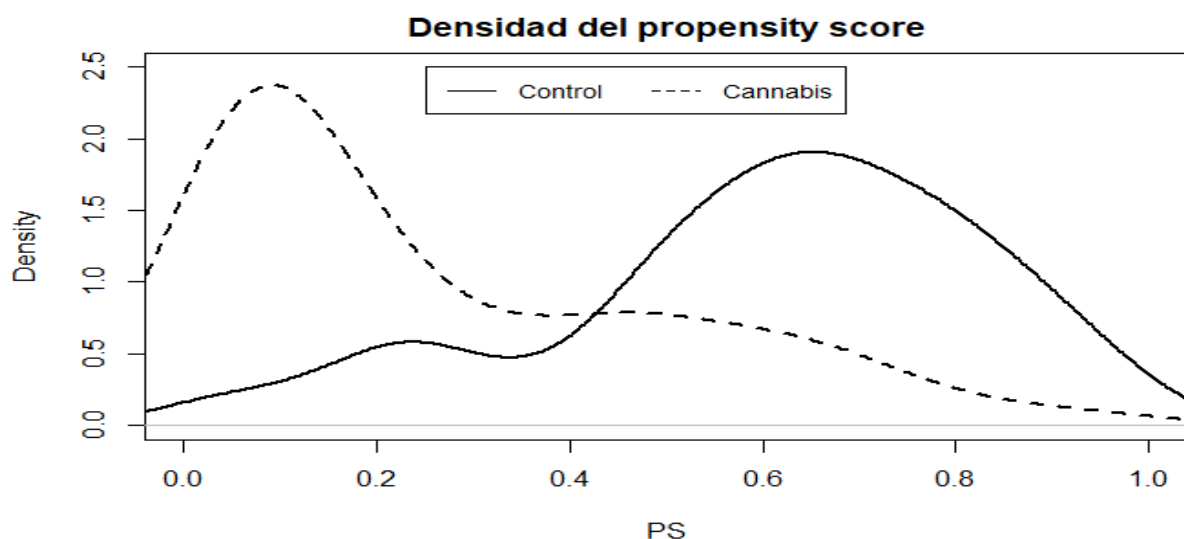
En la Tabla 8 se presenta los valores del propensity score estimado utilizando la función `pscore()`.

Tabla 8 Los cinco primeros valores del propensity score estimado utilizando la función `pscore()`

	pscore	group
1	0.1073165	0
2	0.1634825	0
3	0.7642307	1
4	0.7276919	1
5	0.5284793	0

En la Figura 11 se presenta la ilustración de la densidad del propensity score del grupo cannabis y grupo control. Los individuos del grupo cannabis presentaron una media de 0.26 (0.24), inferior a la que presentan los individuos del grupo control 0.60 (0.22).

Figura 11 Estimación de la densidad del propensity score de los individuos del grupo cannabis y grupo control utilizando la función `plot.pscore()`

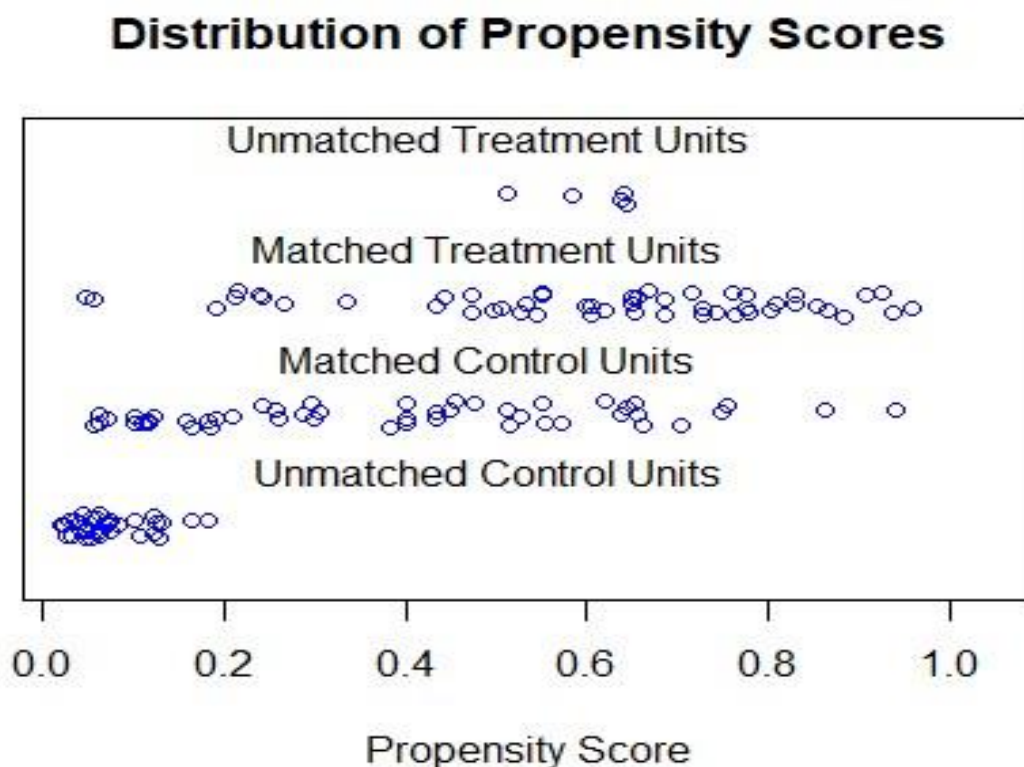


A continuación, presentamos, el emparejamiento por el propensity score para la generación de una muestra sintética emparejado.

4.2.3. Matching por el propensity score

El propensity score tiene varios usos, una de ellas es para la producción de muestras emparejadas. Aplicamos a nuestro estudio para realizar el emparejamiento de los individuos del grupo Cannabis vs grupo Control aplicando la técnica del vecino más cercano dentro de una distancia de caliper. Los individuos del grupo Cannabis se comparan con los individuos del grupo Control en una proporción de uno a uno (emparejamiento 1:1) utilizando el tamaño de caliper predeterminado. La configuración predeterminada para el tamaño de caliper es una quinta parte (0.2) de la desviación estándar del logit del propensity score estimado. La variable del emparejamiento se estableció en el propensity score estimado. En la Figura 12 se presenta una ilustración de la distribución del propensity score estimado después del matching y (Salida de R 1) resume la información completa del matching por el propensity score.

Figura 12: Distribución de puntajes de propensión de unidades de tratamiento (individuos del grupo cannabis) y unidades de control (individuos del grupo control) emparejados y no emparejados utilizando la función `plot()`



Se observa que el emparejamiento no es completo, en el sentido de que hay individuos que no han sido emparejados del grupo de tratamiento y grupo control. La información contenida en la (salida de R 1) detalla lo que se observa en la Figura 12.

Salida de R 1: Matching por el propensity score mediante el método del vecino más cercano dentro de una distancia de caliper utilizando la función `ps.match()`

Matched by: pscore

Matching parameter:

Caliper size: 0.328

Ratio: 1.000

Who is treated?: 1.000

Matching information:

Untreated to treated?: TRUE

Best match?: TRUE

Matching data:

Number of treated obs.: 56

Number of matched treated obs.: 51

Number of untreated obs.: 85

Number of matched untreated obs.: 51

Number of total matched obs.: 102

Number of not matched obs.: 39

Number of matching sets: 51

Number of incomplete matching sets: 0

El parámetro del matching (caliper size: 0.328) es el resultado de la quinta parte (0.2) de la desviación estándar del logit del propensity score estimado. Este parámetro indica que el matching mediante la técnica del vecino más cercano se ha procesado limitando la distancia máxima por 0.328. Mediante la técnica del vecino más cercano dentro de una distancia de caliper se han emparejado 102 individuos en total, 51 individuos del grupo Cannabis y 51 individuos del grupo Control, quedando sin emparejar 5 individuos del grupo Control y 34 individuos del grupo Cannabis, en total quedando sin emparejar 39 individuos.

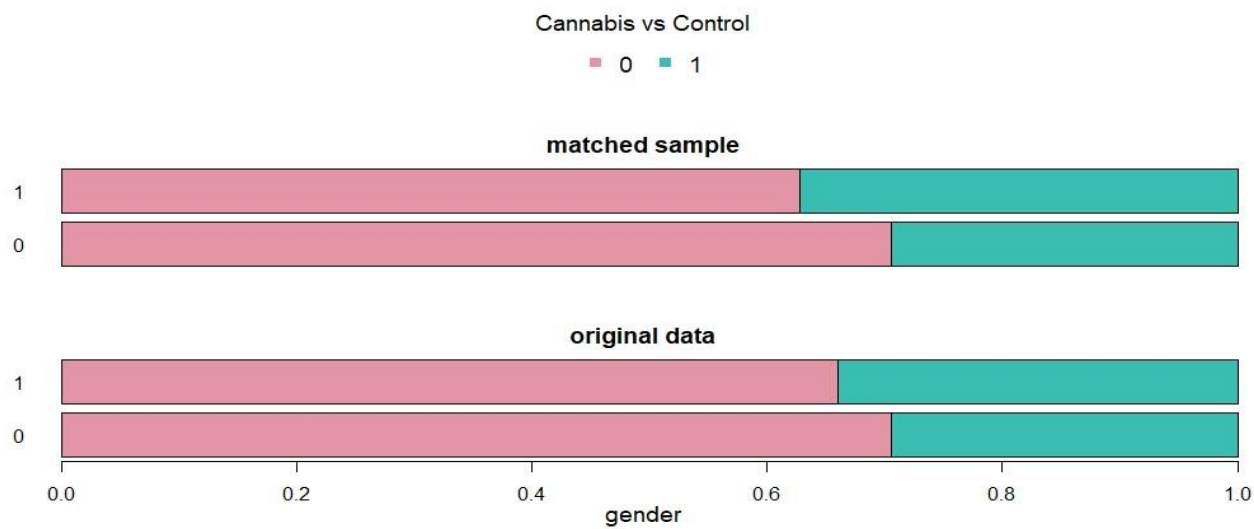
4.2.4. Verificación de equilibrio en la distribución de covariables

Después del emparejamiento una cuestión importante es la verificación del equilibrio de la distribución de las covariables entre los grupos de Cannabis y Control. Para la verificación de equilibrio aplicamos gráficos, pruebas estadísticas y diferencias estandarizadas. La verificación de equilibrio asegura la comparación de usuarios con o sin problemas de cannabis que son lo más similares posible en las covariables medidas.

4.2.5. Verificación de equilibrio mediante gráficos

La verificación de equilibrio para las covariables se realiza antes y después del matching usando los gráficos de barras apiladas para mostrar frecuencias (o más bien proporciones) para covariables categóricas y para las covariables continuas mediante histogramas. En la Figura 13 se presenta una ilustración para el covariable género (gender) y en la salida de R 2 los resultados de las proporciones.

Figura 13 Verificación de equilibrio para la covariable categórica “gender” en los datos antes y después de realizar el matching utilizando la función `dist.plot()`



Filas: categorías de “gender” (0: Hombre, 1: Mujer)
Columnas: categorías de “group” (0: Cannabis, 1: Control)

Salida de R 2: Proporciones para la covariable “gender” (filas) según el tratamiento (columnas) en los datos originales y para la muestra emparejada

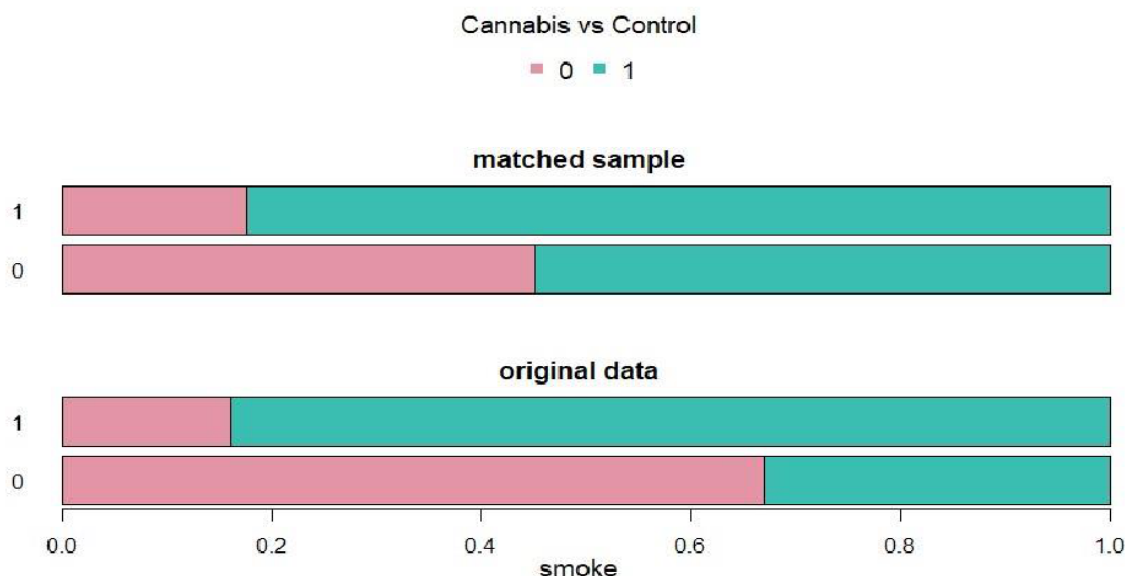
```
[1] "gender"
[[1]]
, , index = original data      , , index = matched sample

  treat      0      1      treat      0      1
0 0.7058824 0.6607143 0 0.7058824 0.6274510
1 0.2941176 0.3392857 1 0.2941176 0.3725490
```

Filas: categorías de “gender” (0: Hombre, 1: Mujer)
Columnas: categorías de “group” (0: Cannabis, 1: Control)

No existe una diferencia de proporción en el grupo Control (etiquetado por ‘0’), pero si existe una diferencia de proporciones en los individuos del grupo Cannabis, ya que los individuos del grupo Cannabis con los Controles fueron emparejados). Antes y después del emparejamiento las proporciones de “gender” solo experimentan cambios muy pequeños. A continuación, presentamos la verificación de equilibrio para la covariable estado de fumar “smoke” mediante una ilustración gráfica (Figura 14) y la salida de R 3 nos da los resultados de las proporciones.

Figura 14 Verificación de equilibrio para la covariable categórica “smoke” antes y después de realizar el matching utilizando la función `dist.plot()`



Filas: categorías de “smoke” (0: Current smoker, 1: Non smoker/Exsmoker).

Columnas: categorías de “group” (0: Cannabis, 1: Control).

Salida de R 3: Proporciones de la covariable “smoke” (filas) según el grupo de tratamiento (columnas) en los datos originales y para la muestra emparejada

```
[1] "smoke"
[[1]]
, , index = original data      , , index = matched sample

      treat      treat
      0      1      0      1
0  0.6705882 0.1607143 0.4509804 0.1764706
1  0.3294118 0.8392857 0.5490196 0.8235294
2
```

Filas: categorías de “smoke” (0: Current smoker, 1: Non smoker/Exsmoker).

Columnas: categorías de “group” (0: Cannabis, 1: Control).

Para la covariable “smoke” existe una diferencia de proporción en el grupo Control (etiquetado por “0”). También existe una diferencia de proporción en el grupo Cannabis (etiquetado por “1”). Después del matching (“muestra emparejada”), las proporciones de “smoke” son más similares en ambos grupos en comparación con los datos originales. A continuación, se realiza un análisis gráfico para verificar el equilibrio para las covariables continuas antes y después del emparejamiento. La distribución de covariables continuas “age” y “civoc” se presenta mediante histogramas, Figura 15 y Figura 16, con sus respectivas tablas de frecuencias (Tabla 9 y Tabla 10).

Figura 15: Verificación de equilibrio mediante histogramas de la covariable edad (en años) antes y después del matching utilizando la función `dist.plot()`

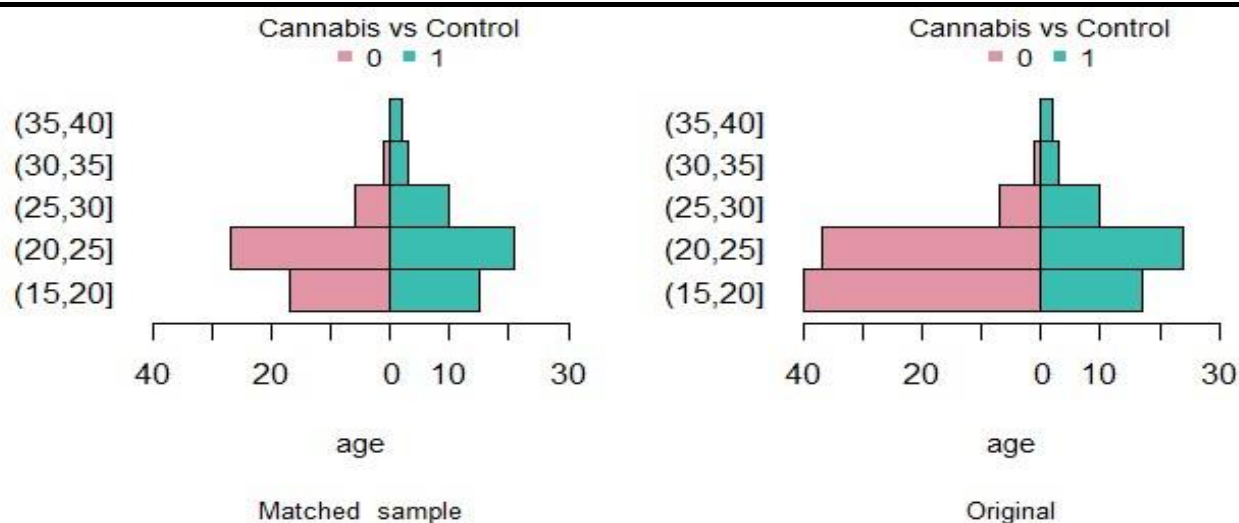


Tabla 9 Distribución de frecuencias de la covariable edad según los grupos de tratamiento para datos originales y muestra emparejada

Edad	Cannabis=0		Control=1	
	Datos originales	Muestra emparejada	Datos originales	Muestra emparejada
(15, 20]	40	17	17	15
(20, 25]	37	27	24	21
(25, 30]	7	6	10	10
(30, 35]	1	1	3	3
(35, 40]	0	0	2	2

En el grupo Cannabis (datos originales) no hay individuos que tenga edad en el intervalo (35, 40] para realizar el emparejamiento, en consecuencia, observamos que en la muestra emparejada no existen individuos emparejados en el intervalo (35, 40].

Figura 16 Verificación de equilibrio de la covariable *civoc* antes y después del matching utilizando la función `dist.plot()`

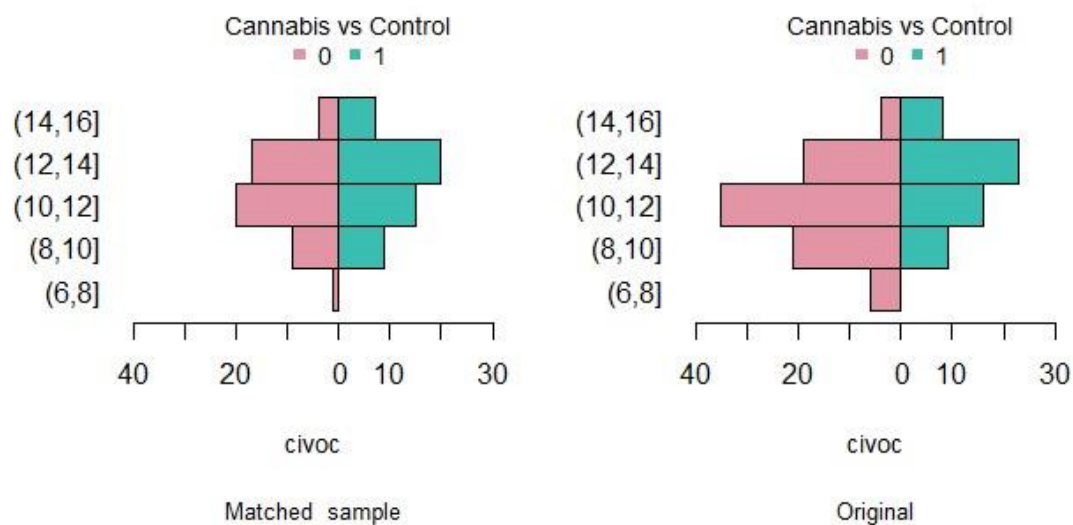


Tabla 10 Distribución de frecuencias de la covariable *civoc* según los grupos de tratamiento antes y después del matching

Civoc	Cannabis=0		Control=1	
	Datos originales	Muestra emparejada	Datos originales	Muestra emparejada
(6, 8]	6	1	0	0
(8, 10]	21	9	9	9
(10, 12]	35	20	16	15
(12, 14]	19	17	23	20
(14, 16]	4	4	8	7

Para la covariable “*civoc*” en el grupo Cannabis se ha procesado el matching en los cinco niveles del intervalo, sin embargo, en el grupo control, en el primer intervalo no hay individuos emparejados. La verificación de equilibrio por medio de gráficos nos proporciona las primeras ideas intuitivas de equilibrio.

4.2.6. Verificación de equilibrio con pruebas estadísticas

La ilustración grafica de las distribuciones de covariables en los grupos de tratamiento da una primera idea, sin embargo, las pruebas estadísticas deciden si las diferencias potencialmente visibles en los gráficos son significativas. Cuatro covariables (gender, age, civoc y smoke) en la muestra emparejada son verificadas por pruebas estadísticas para el equilibrio entre los grupos de tratamiento. A continuación, se presenta salida de R 4 que contiene resultados de pruebas estadísticas antes y después de realizar el emparejamiento, para decidir sobre el equilibrio en la muestra emparejada.

Salida de R 4: Verificación de equilibrio mediante pruebas estadísticas antes y después del emparejamiento utilizando la función `ps.balance()`

```
> summary(cannabi.m1.bal)
```

```
Balance check using: Statistical tests
```

```
Summary of balance check:
```

	Before: no bal (0)	Before: bal (1)
	-----	-----
After: no bal (0)	1	0

After: bal (1)	2	1

```
Covariates not completely tested: ---
```

```
Detailed balance check (overall):
```

	gender	age	civoc	smoke
Before	1	0	0	0
-----	-----	-----	-----	-----
After	1	1	1	0

```
Detailed balance check [p.values]:
```

	gender	age	civoc	smoke
Before	0.703	0.019	0	0
-----	-----	-----	-----	-----
After	0.529	0.111	0.286	0.006
-----	-----	-----	-----	-----
Test	chi^2	t	t	chi^2

```
Significance level for tests: 0.05
```

En el resultado, hay cuatro informaciones diferentes disponibles. Primero, una tabla resumen para la decisión de equilibrio o sin equilibrio. Hay una covariable balanceada y tres no balanceadas antes del emparejamiento (columnas de la tabla “Summary of balance check”). Después del emparejamiento tenemos tres covariables balanceadas y una covariable no balanceada (filas de la tabla “Summary of balance check”). La siguiente información (“Covariates not completely tested”) es sobre covariables para las cuales las pruebas en los datos originales o muestra emparejada no pudieron ser realizado correctamente, para nuestro estudio indica que se hicieron las pruebas para las cuatro covariables.

La tabla “Detailed balance check (overall)” presenta información de equilibrio específica de covariable, donde, “0” indica “no equilibrado” y “1” significa “equilibrado”. Es notable que la covariable “género” se equilibra (indicado por “1”) antes y después del emparejamiento. Las covariables “age” y “civoc” (“1”) también se equilibran después del emparejamiento y la covariable “smoke” (“0”) permanece desequilibrada. La cuarta y la última información muestran los p-valores de las pruebas estadísticas aplicadas en los datos originales y muestra emparejada. La covariable “smoke” es el único que tiene un p-valor inferior al nivel de significancia (0.05).

4.2.7. Verificación de equilibrio con diferencias estandarizadas

El equilibrio de covariables “gender”, “age”, “civoc” y “smoke” entre los grupos de tratamiento es de interés antes y después del emparejamiento. Por lo tanto, las diferencias estandarizadas son computados en base a las fórmulas descritas en el documento. A continuación, se presenta salida de R 5 de la verificación de equilibrio mediante diferencias estandarizadas.

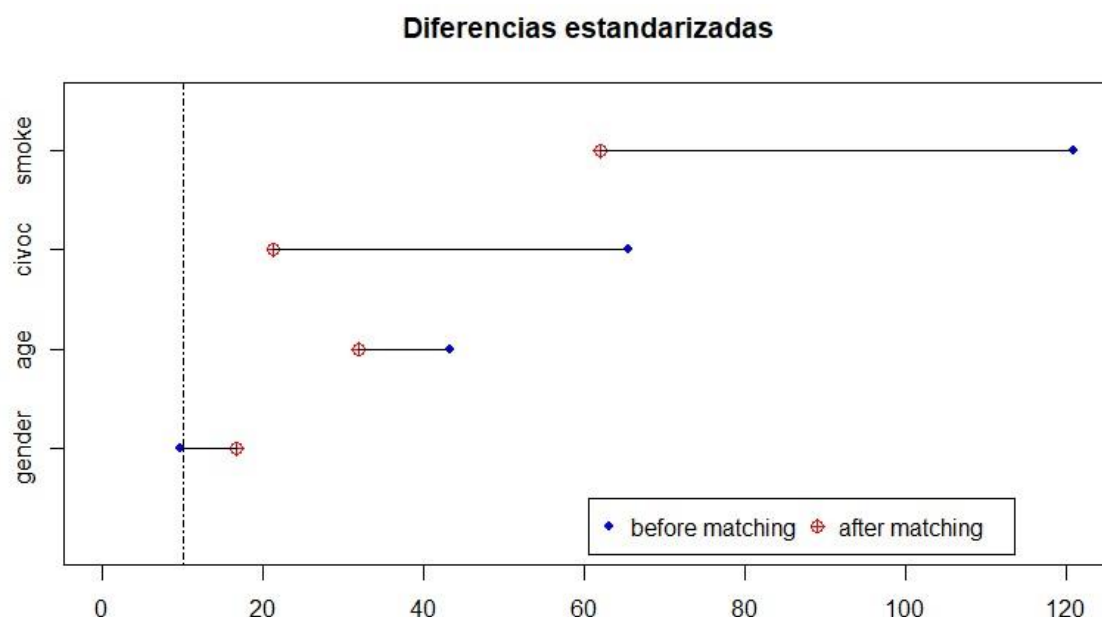
Salida de R 5: Verificación de equilibrio mediante diferencias estandarizadas antes y después del emparejamiento utilizando la función `ps.balance()`

```
> cannabi.m1.bal_est
Detailed balance check:
[Standardized differences (cut point: 10)]
```

	gender	age	civoc	smoke
Before	9.721	43.277	65.527	120.888
-----	-----	-----	-----	-----
After	16.696	31.848	21.222	61.932
-----	-----	-----	-----	-----
Scale	bin	num	num	bin

Con el matching se consiguió una reducción del valor de las diferencias para las tres covariables (“age”, “civoc” y “smoke”), pero las diferencias continúan siendo significantes, sin embargo, antes del emparejamiento hay una covariable “gender” con diferencia estandarizada insignificante. Se consideran diferencias insignificantes valores <10%. En la Figura 17 se puede contemplar gráficamente la idea de las diferencias estandarizadas.

Figura 17 Una ilustración de las diferencias estandarizadas antes y después del matching utilizando la función `plot.stdf()`



4.2.8. Ajuste de la muestra emparejada en el modelo ANCOVA

Hasta aquí, ya tenemos las covariables incluidas en el modelo PS suficientemente equilibradas después del emparejamiento. Con la muestra emparejada se procede con el ajuste de las covariables en el modelo ANCOVA ($\text{var} \sim \text{group} + \text{gender} + \text{age} + \text{civoc} + \text{smoke}$), obteniendo los resultados para comparar los consumidores de cannabis y los controles respecto al rendimiento en las distintas variables neuropsicológicas. En la Tabla 11 se presenta los resultados obtenidos mediante el método de emparejamiento por el propensity score.

Tabla 11 Rendimiento Neuropsicológico en Función del Estado de Consumo de Drogas: Usuarios de cannabis vs Controles

	Probability of hit		Probability of hit (N=2)		Total errors adjusted	
	Modelo ANCOVA	Emparejamiento por PS	Modelo ANCOVA	Emparejamiento por PS	Modelo ANCOVA	Emparejamiento por PS
Cannabis (85)						
Mean	67.5	69.2	90.7	92.5	25.9	24.1
SD	17.3	18.2	10.72	5.79	21.9	21
Control (56)						
Mean	72.5	71.5	93.4	93.8	22.8	23.1
SD	13.8	13.5	8.98	8.69	19.9	20
Group F-value	0.015	0.003	0.174	0.011	0.066	0.114
P-value	0.901	0.959	0.677	0.916	0.798	0.736
	Mean latency to first choice		Net score		Z composite score	
	Modelo ANCOVA	Emparejamiento por PS	Modelo ANCOVA	Emparejamiento por PS	Modelo ANCOVA	Emparejamiento por PS
Cannabis (85)						
Mean	14184	14004	4.56	1.22	-0.482	-0.263
SD	4928	4685	21.5	20.1	2.97	3.03
Control (56)						
Mean	16268	16327	10.84	8.80	0.659	0.478
SD	5899	6075	24.3	23.2	2.98	2.81
Group F-value	2.805	2.793	3.549	3.311	1.004	0.934
P-value	0.096	0.098	0.062	0.072	0.318	0.336

Los resultados se presentan como media (SD).

Los modelos ANCOVA se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Emparejamiento por PS se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Comparaciones grupales

Con el método del matching por el propensity score no encontramos diferencias significativas entre los usuarios de cannabis y los controles en ninguno de los seis principales índices de rendimiento de las tareas. Solo se observa cambios leves en los resultados numéricos comparados con los resultados del modelo ANCOVA.

Los resultados numéricos en general serán aproximadamente casi iguales cuando los estudios se hacen con muestras grandes, es importante señalar que este estudio se ha realizado solo con una muestra de 141 observaciones, también es de destacar que cuando se hace el emparejamiento el tamaño de la muestra original se reduce y eso tienen implicancias en que los resultados varíen levemente. Este es el método preferido y más utilizado ya que al final se llega a obtener las mismas conclusiones en contraste con los modelos ANCOVA.

4.2.9. Estratificación por el propensity score

Este técnico agrupa a individuos que tienen valores similares o iguales usando el propensity score como variable de estratificación. A continuación, se presenta salida de R 6 que contiene la información de la proporción de individuos en cada uno de los quintiles del propensity score.

Salida de R 6: Estratificación por quintiles del propensity score utilizando la función `ps.makestrata()`

```
> summary.stratified.pscore(cannabi.str5)
Stratified by:  pscore
Strata information:
```

	Strata bounds	n	n (proportion)
	-----	--	-----
1	[0.0207,0.0738]	29	0.206
2	(0.0738,0.242]	29	0.206
3	(0.242,0.512]	27	0.191
4	(0.512,0.662]	28	0.199
5	(0.662,0.959]	28	0.199

Para muestras pequeñas se recomienda estratificar usando quintiles del propensity score y para muestras grandes es preferible estratificar mediante el uso de los deciles del propensity score. A continuación, se presenta salida de R 7 con los resultados de la estratificación mediante los deciles del propensity score.

Salida de R 7: Estratificación por deciles del propensity score utilizando la función `ps.makestrata()`

`> summary.stratified.pscore(cannabi.str10)`

Stratified by: pscore

Strata information:

	Strata bounds	n	n (proportion)
	-----	--	-----
1	[0.0207,0.0551]	15	0.106
2	(0.0551,0.0738]	14	0.099
3	(0.0738,0.129]	14	0.099
4	(0.129,0.242]	15	0.106
5	(0.242,0.401]	13	0.092
6	(0.401,0.512]	14	0.099
7	(0.512,0.605]	14	0.099
8	(0.605,0.662]	14	0.099
9	(0.662,0.775]	14	0.099
10	(0.775,0.959]	14	0.099

Ya que disponemos de una muestra relativamente pequeño, con 141 observaciones, tomamos la decisión de trabajar con la muestra estratificada por quintiles del propensity score. Después de la estratificación, se procede con la verificación de equilibrio analizando las distribuciones de las covariables entre los grupos de Cannabis vs Control.

4.2.9.1. Verificación de equilibrio mediante gráficos

La verificación de equilibrio para la muestra estratificada se hace de manera similar caso muestra emparejada. A continuación, se presenta un gráfico para la covariable “smoke” (Figura 18), ya que, es la covariable que no se equilibra suficientemente analizado en la verificación de equilibrio según el emparejamiento por el propensity score. Adicionalmente se presenta la Tabla12 que contiene las proporciones y medias antes y después de la estratificación.

Figura 18 Verificación de equilibrio para la covariable “smoke” antes y después de la estratificación utilizando la función `dist.plot()`

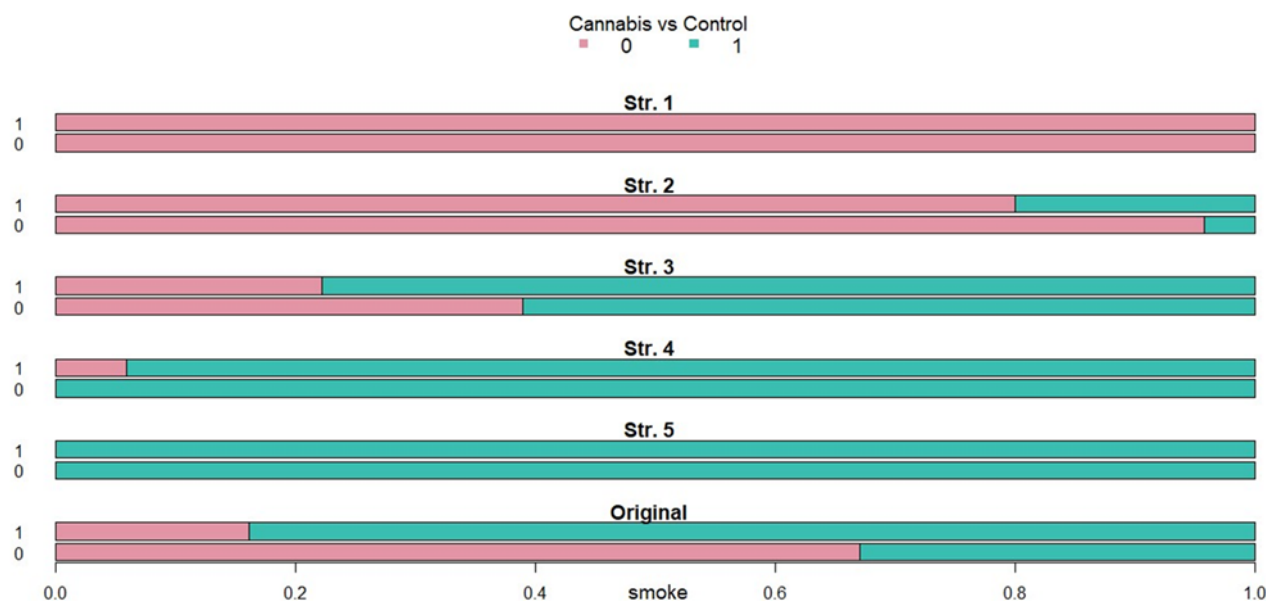


Tabla 12 Verificación de equilibrio mediante la comparación de medias para covariables continuas y mediante proporciones para las covariables categóricas antes y después de la estratificación

	Str. 1		Str. 2		Str. 3		Str. 4		Str. 5		Datos Originales	
Usuarios	(27)	(2)	(24)	(5)	(18)	(9)	(11)	(17)	(5)	(23)	(85)	(56)
Group	0	1	0	1	0	1	0	1	0	1	0	1
age	20.1	20.5	22.3	22.2	21.5	21.2	21.6	22.5	24.8	25.0	21.5	23.2
civoc	9.8	10.0	11.9	12.0	11.7	11.8	11.7	11.9	14.2	13.7	11.3	12-6
Gender												
Male	0.81	1.00	0.71	0.40	0.56	0.78	0.64	0.71	0.80	0.61	0.71	0.66
Female	0.19	0.00	0.29	0.60	0.44	0.22	0.36	0.29	0.20	0.39	0.29	0.34
Smoker												
CS	1.00	1.00	0.96	0.80	0.39	0.22	0.00	0.06	0.00	0.00	0.67	0.16
NS/ES	0.00	0.00	0.04	0.20	0.61	0.78	1.00	0.94	1.00	1.00	0.33	0.84

Abreviaturas: CS y NS/ES son los niveles de smoke

Según la Tabla 12 las medias de la covariable edad y civic tienen una distribución aproximadamente similar antes y después de la estratificación. En las covariables categóricas las variaciones de las proporciones son más acentuadas. En el estrato 1 para la covariable smoker todos los 27 individuos del grupo Cannabis son de la categoría “Current smoker” y los dos individuos del grupo control son también de la categoría “Current smoker”, es decir no hay individuos en la categoría “Non-smoker/Exsmoker”, similarmente sucede en el estrato 5. Este hecho tendrá implicancias cuando se hace prueba estadística.

4.2.9.2. Verificación de equilibrio con pruebas estadísticas

En la salida de R 8 se presenta los resultados de las pruebas estadísticas: resumen de la verificación de equilibrio, covariables no completamente probadas, verificación de equilibrio detallada general) y los resultados de la prueba, es decir, p-valores de las pruebas estadísticas aplicadas en los datos originales y en la muestra estratificado.

Salida de R 8: Verificación de equilibrio mediante pruebas estadísticas antes y después de la estratificación utilizando la función `ps.balance()`

```
> summary(cannabi.str5.bal)
Balance check using: Statistical tests
Summary of balance check:
               Before: no bal (0)   Before: bal (1)
               -----
After: no bal (0)   0
-----
After: bal (1)     2
-----

Covariates not completely tested: smoke
Detailed balance check (overall):
Before    gender    age    civic    smoke
-----
After     1        1      1        NA
-----
Detailed balance check (per stratum) [p.values]:
Before    gender    age    civic    smoke
-----
Stratum 1  1        0.902  0.862  NA
Stratum 2  0.422    0.98   0.899  0.763
Stratum 3  0.481    0.855  0.937  0.665
Stratum 4  1        0.496  0.72   1
Stratum 5  0.769    0.913  0.549  NA
-----
Test       chi^2     t      t      chi^2
```

Significance level for tests: 0.05

Las pruebas para tres covariables (“gender”, “age” y “civoc”) fueron realizado correctamente y para la covariable “smoke” no pudieron ser realizado correctamente en los estratos 1 y 5. Hay una covariable balanceado y dos no balanceados antes de la estratificación (columnas de la tabla “Resumen de verificación de equilibrio”). La estratificación resulta en tres covariables balanceadas (filas de la tabla “Resumen de verificación de equilibrio”). La tabla “Verificación de equilibrio detallada (general)” presenta información de equilibrio específica de las covariables donde “0” indica no equilibrado y “1” indica equilibrado. Las covariables “age” y “civoc” se equilibran después de la estratificación y la covariable “gender” permanece equilibrada. Las covariables “gender”, “age” y “civoc” tienen p-valores en cada uno de los cinco estratos superiores al nivel de significación establecido (0.05). Los “NA” significa que no se ha realizado ninguna prueba, es decir, covariable “smoke” no se verifica en los estratos 1 y 5, ya que no hay individuos en los estratos 1 y 5 con “Non- smoker/Ex- smoker =1” y “current smoker=0” respectivamente.

4.2.9.3. Verificación de equilibrio con diferencias estandarizadas

A continuación, se presenta salida de R 9 con los resultados de las diferencias estandarizadas para las cuatro covariables “gender”, “age”, “civoc” y “smoke” en los datos originales y en la muestra estratificada.

Salida de R 9: Verificación de equilibrio mediante diferencias estandarizadas antes y después de la estratificación utilizando la función `ps.balance()`

```
> cannabi.str5.bal_est
```

```
Detailed balance check (per stratum):  
[Standardized differences (cut point: 10)]
```

	gender	age	civoc	smoke
Before	9.721	43.277	65.527	120.888
-----	-----	-----	-----	-----
Stratum 1	67.42	13.497	14.588	NA
Stratum 2	65.25	1.445	7.418	50.078
Stratum 3	48.507	7.453	3.076	36.788
Stratum 4	14.838	24.51	13.459	35.355
Stratum 5	42.874	4.525	33.507	NA
-----	-----	-----	-----	-----
Scale	bin	num	num	bin

Para la covariable “smoke” las diferencias estandarizadas en los estratos 1 y 5 no pudieron ser realizado, ya que no hay individuos en los estratos 1 y 5 con “Non- smoker/Ex- smoker =1” y “current smoker=0” respectivamente. Para la covariable “gender” las diferencias estandarizadas en los cinco estratos son estadísticamente significativas, pero antes de la estratificación resulta insignificante. Por otro lado, las diferencias estandarizadas en los estratos 2, 3 y 5 para la covariable “age” son insignificantes y en los estratos 2 y 3 para la covariable “civoc” también son insignificantes.

4.2.9.4. Ajuste de la muestra estratificada en el modelo ANCOVA

Después de la verificación de equilibrio, ya tenemos la muestra estratificada, para el procesamiento y obtener resultados para comparar los consumidores de cannabis y los controles respecto al rendimiento en las distintas variables neuropsicológicas. Los resultados se reproducen ajustando las covariables de la muestra estratificada en el modelo ANCOVA (var ~ group + gender + age + civoc + smoke). En la Tabla 13 se presenta los resultados obtenidos mediante el método de estratificación por el propensity score.

Tabla 13 Rendimiento Neuropsicológico en Función del Estado del Consumo de Drogas: Usuarios de Cannabis vs Controles

	Probability of hit		Probability of hit (N=2)		Total errors adjusted	
	Modelo ANCOVA	Estratificación por PS	Modelo ANCOVA	Estratificación por PS	Modelo ANCOVA	Estratificación por PS
Cannabis (85)						
Mean	67.5	67.5	90.7	92.5	25.9	25.9
SD	17.3	17.3	10.72	5.79	21.9	21.9
Control (56)						
Mean	72.5	72.5	93.4	93.8	22.8	22.8
SD	13.8	13.8	8.98	8.69	19.9	19.9
Group						
F-value	0.015	0.149	0.174	0.148	0.066	0.872
P-value	0.901	0.700	0.677	0.701	0.798	0.352
	Mean latency to first choice		Net score		Z composite score	
	Modelo ANCOVA	Estratificación por PS	Modelo ANCOVA	Estratificación por PS	Modelo ANCOVA	Estratificación por PS
Cannabis (85)						
Mean	14184	14184	4.56	4.56	-0.482	-0.482
SD	4928	4928	21.5	21.5	2.97	2.97
Control (56)						
Mean	16268	16268	10.84	10.84	0.659	0.659
SD	5899	5899	24.3	24.3	2.98	2.98
Group						
F-value	2.805	4.454	3.549	6.884	1.004	1.949
P-value	0.096	0.037	0.062	0.010	0.318	0.165

Los resultados se presentan como media (SD).

Los modelos ANCOVA se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Estratificación por PS se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Comparaciones grupales

En contraste con los modelos ANCOVA, mediante el método de estratificación por el propensity score, hemos encontrado dos diferencias significativas de los seis principales índices de rendimiento. Primero, hay una diferencia significativa entre los grupos en Mean latency to first choice, que se considera que indexa la “planificación de tareas”. Segundo, hemos encontrado diferencias significativas entre los usuarios de cannabis y los controles en Net score, que indexa “calidad de toma de decisiones”. Estos posibles cambios en el nivel de significancia en los resultados podrían haberse derivado, primero de la insuficiencia de balanceo de la covariable “smoke” y otra posibilidad es haber realizado el estudio con una muestra relativamente pequeña, que cuando se estratifica por los quintiles del propensity score, se pierde observaciones en los grupos y niveles de las covariables en algunos estratos. La estratificación sería ideal cuando se dispone de muestras grandes, pero no complejas. Para 141 observaciones, mediante este método no podemos asegurar que los cambios en los resultados encontrados en contraste con los del modelo ANCOVA estén garantizados.

4.2.10. Ponderación por el propensity score

El método de ponderación por el propensity score consiste en ponderar a los individuos del grupo Cannabis por la inversa del propensity score ($1/PS$), mientras a los individuos del grupo control se le asigna un peso igual a ($1/PS - 1$). La probabilidad inversa de ponderación utilizando el puntaje de propensión, usa pesos basados en el puntaje de propensión para crear una muestra sintética en la que la distribución de las covariables medidas es independiente de la asignación al tratamiento, Austin (2011). Mediante esta técnica ya no realizamos verificación de equilibrio, ya que, después de la ponderación se obtiene una muestra sintética con la distribución de las covariables medidas independiente de la asignación al tratamiento. A continuación, para la obtención de resultados, presentamos el ajuste de la muestra ponderada en el modelo ANCOVA.

4.2.11. Ajuste de la muestra ponderada en el modelo ANCOVA

Con la muestra ponderada por el propensity score procedemos con el ajuste de las covariables en el modelo ANCOVA ($\text{var} \sim \text{group} + \text{PS}$). En la Tabla 14 se presenta los resultados reproducidos mediante este método para comparar los consumidores de Cannabis y los Controles respecto al rendimiento cognitivo en las distintas variables neuropsicológicas.

Tabla 14 Rendimiento Neuropsicológico en Función del Estado de Consumo de Drogas: Usuarios de cannabis vs Controles

	Probability of hit		Probability of hit (N=2)		Total errors adjusted	
	Modelo ANCOVA	Ponderación por PS	Modelo ANCOVA	Ponderación por PS	Modelo ANCOVA	Ponderación por PS
Cannabis (85)						
Mean	67.5	67.5	90.7	90.7	25.9	25.9
SD	17.3	17.3	10.72	10.72	21.9	21.9
Control (56)						
Mean	72.5	72.5	93.4	93.4	22.8	22.8
SD	13.8	13.8	8.98	8.98	19.9	19.9
Group						
F-value	0.015	0.105	0.174	0.204	0.066	0.049
P-value	0.901	0.746	0.677	0.652	0.798	0.825
	Mean latency to firts choice		Net score		Z composite score	
	Modelo ANCOVA	Ponderación por PS	Modelo ANCOVA	Ponderación por PS	Modelo ANCOVA	Ponderación por PS
Cannabis (85)						
Mean	14184	14184	4.56	4.56	-0.482	-0.482
SD	4928	4928	21.5	21.5	2.97	2.97
Control (56)						
Mean	16268	16268	10.84	10.84	0.659	0.659
SD	5899	5899	24.3	24.3	2.98	2.98
Group						
F-value	2.805	0.341	3.549	5.141	1.004	2.442
P-value	0.096	0.560	0.062	0.025	0.318	0.120

Los resultados se presentan como media (SD).

Los modelos ANCOVA se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Ponderación por PS se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Comparaciones grupales

Mediante el método de ponderación por el propensity score no observa diferencias significativas entre los usuarios de cannabis y los controles en ninguno de los principales índices de rendimiento de las tareas. Por lo tanto, el modelo ajustado por el propensity score no produce cambios en los resultados en contraste con el modelo ANCOVA, ya que, el p-valor del estadístico de Hosmer-Lemeshow $> 0,05$ en ambos modelos.

4.2.12. El propensity score como variable de ajuste en el modelo ANCOVA

Aplicamos este método para comparar los consumidores de cannabis y los controles respecto al rendimiento en las distintas variables neuropsicológicas usando el propensity score como variable de ajuste. Para reproducir los resultados, el propensity score y el grupo de estudio se incluyen en el modelo ANCOVA como variables de ajuste y las distintas variables neuropsicológicas ($\text{var} \sim \text{group} + \text{PS}$). En la Tabla 15 se presentan los resultados del análisis respecto al rendimiento en las distintas variables neuropsicológicas usando el propensity score como variable de ajuste y los resultados del modelo ANCOVA ajustados como ($\text{var} \sim \text{group} + \text{gender} + \text{age} + \text{civoc} + \text{smoke}$).

Tabla 15 Rendimiento Neuropsicológico en Función del Estado de Consumo de Drogas: Usuarios de cannabis vs Controles

	Probability of hit		Probability of hit (N=2)		Total errors adjusted	
	Modelo ANCOVA	Ajuste por PS	Modelo ANCOVA	Ajuste por PS	Modelo ANCOVA	Ajuste por PS
Cannabis (85)						
Mean	67.5	67.5	90.7	90.7	25.9	25.9
SD	17.3	17.3	10.72	10.72	21.9	21.9
Control (56)						
Mean	72.5	72.5	93.4	93.4	22.8	22.8
SD	13.8	13.8	8.98	8.98	19.9	19.9
Group						
F-value	0.015	0.003	0.174	0.158	0.066	0.132
P-value	0.901	0.953	0.677	0.691	0.798	0.717
	Mean latency to firts choise		Net score		Z composite score	
	Modelo ANCOVA	Ajuste por PS	Modelo ANCOVA	Ajuste por PS	Modelo ANCOVA	Ajuste por PS
Cannabis (85)						
Mean	14184	14184	4.56	4.56	-0.482	-0.482
SD	4928	4928	21.5	21.5	2.97	2.97
Control (56)						
Mean	16268	16268	10.84	10.84	0.659	0.659
SD	5899	5899	24.3	24.3	2.98	2.98
Group						
F-value	2.805	2.383	3.549	2.811	1.004	0.825
P-value	0.096	0.125	0.062	0.096	0.318	0.365

Los resultados se presentan como media (SD).

Los modelos ANCOVA se ajustan por sexo, edad, tabaquismo y coeficiente intelectual.

Ajuste por PS se ajustan por grupo y el propensity score.

Comparaciones grupales

Mediante este enfoque revolucionario observamos que no hay diferencias significativas entre los usuarios de cannabis y los controles en ninguno de los seis principales índices de rendimiento de las tareas. Por lo tanto, el modelo ajustado por el propensity score no produce cambios en los resultados en contraste con el modelo ANCOVA, ya que, el p-valor del estadístico de Hosmer-Lemeshow $> 0,05$ en ambos modelos. Este método, nos da mejores resultados en cuanto a la calidad de resultados numéricos idénticos comparados con los del modelo ANCOVA. El único inconveniente de esta metodología es que requiere que se especifique un modelo para poder realizar el análisis. Generalmente este método da resultados similares e iguales en muchos estudios que se han publicado en la literatura. Y su versatilidad de reducir la dimensionalidad, en el sentido de que en vez de ajustar por 100 covariables ajustar únicamente por uno solo (el propensity score) es uno de los avances más importantes para realizar análisis con datos observacionales. A continuación, presentamos una tabla general (Tabla 16) que contiene los resultados de los cuatro métodos del propensity score, que incluye los resultados obtenidos con el ajuste clásico de modelos ANCOVA. Los resultados de la Tabla 16, serán discutidos, comparados y analizados.

Tabla 16 Resultados globales de los cuatro métodos del propensity score incluido los resultados del modelo ANCOVA

	Modelo ANCOVA	Emparejamiento por PS	Estratificación por PS	Ponderación por PS	Ajuste por PS	Modelo ANCOVA	Emparejamiento por PS	Estratificación por PS	Ponderación por PS	Ajuste por PS
	RVP: Probability of hit					Nback 2: (Hits/30 - FA/70)*100				
Cannabis										
Mean	67.5	69.2	67.5	67.5	67.5	90.7	92.5	90.7	90.7	90.7
SD	17.3	18.2	17.3	17.3	17.3	10.72	5.79	10.72	10.72	10.72
Control										
Mean	72.5	71.5	72.5	72.5	72.5	93.4	93.8	93.4	93.4	93.4
SD	13.8	13.5	13.8	13.8	13.8	8.98	8.69	8.98	8.98	8.98
F-value	0.015	0.003	0.149	0.105	0.003	0.174	0.011	0.148	0.204	0.158
P-value	0.901	0.959	0.700	0.746	0.953	0.677	0.916	0.701	0.652	0.691
	IED Total Errors Adjusted					OTS Mean Latency to First Choice				
Cannabis										
Mean	25.9	24.1	25.9	25.9	25.9	14184	14004	14184	14184	14184
SD	21.9	21	21.9	21.9	21.9	4928	4685	4928	4928	4928
Control										
Mean	22.8	23.1	22.8	22.8	22.8	16268	16327	16268	16268	16268
SD	19.9	20	19.9	19.9	19.9	5899	6075	5899	5899	5899
F-value	0.066	0.114	0.872	0.049	0.132	2.805	2.793	4.454	0.341	2.383
P-value	0.798	0.736	0.352	0.825	0.717	0.096	0.098	0.037	0.560	0.125
	Iowa Gambling Test					Cognition Index				
Cannabis										
Mean	4.56	1.22	4.56	4.56	4.56	-0.482	-0.263	-0.482	-0.482	-0.482
SD	21.5	20.1	21.5	21.5	21.5	2.97	3.03	2.97	2.97	2.97
Control										
Mean	10.84	8.80	10.84	10.84	10.84	0.659	0.478	0.659	0.659	0.659
SD	24.3	23.2	24.3	24.3	24.3	2.98	2.81	2.98	2.98	2.98
F-value	3.549	3.311	6.884	5.141	2.811	1.004	0.934	1.949	2.442	0.825
P-value	0.062	0.072	0.010	0.025	0.096	0.318	0.336	0.165	0.120	0.365

5. DISCUSIÓN Y CONCLUSIONES

5.1. DISCUSIÓN

Con respecto al primer hallazgo, la ausencia de efectos adversos generales del consumo de cannabis en las funciones ejecutivas es consistente con los resultados de los metaanálisis más exhaustivos realizados hasta ahora sobre el vínculo entre el cannabis y la cognición (Grant et al, 2003). Más concretamente, como se observa en la Tabla 19 (resultados globales de los cuatro métodos del propensity score incluida los modelos ANCOVA) al estudiar una muestra de jóvenes consumidores de cannabis frente a los controles que no usan drogas, encontramos asociaciones esencialmente nulas que vinculan el consumo de cannabis con los déficits del rendimiento cognitivo en los jóvenes. En los metaanálisis realizados sobre la relación entre el uso de cannabis y la cognición, se ha llegado a la conclusión de que el principal efecto adverso del consumo de cannabis en la función cognitiva se limita a los dominios de aprendizaje y la memoria, mientras que los efectos no fueron significativos para el dominio de atención y control ejecutivo, que abordamos aquí. Hay estudios realizados que sostienen que los efectos adversos sobre el consumo de cannabis en la función ejecutiva pueden tener relación con la dosis.

Con respecto al tercer hallazgo, se descubrió que la magnitud de la asociación era menor para los jóvenes del grupo control que no usan drogas que para los jóvenes del grupo cannabis, en dos de las funciones ejecutivas, pero con poca evidencia de asociaciones estadísticamente sólidas tanto para el grupo cannabis como para el grupo control. Investigaciones previas que examinaron preguntas similares sugieren que las jóvenes consumidoras de altas dosis de cannabis podrían ser un poco más propensas que los jóvenes del grupo control a experimentar déficits neurocognitivos.

Nuestro tamaño de muestra (n=141) relativamente pequeño puede ser responsable de los resultados cualitativamente diferentes entre usuarios de cannabis y los controles. Lo que si queda claro es que existe una clara necesidad de replicación del estudio con muestras de mayor tamaño. Nuestro estudio se ha enfocado solo en la evaluación de algunos aspectos, no respalda la hipótesis de los polimorfismos genéticos en los efectos adversos en el rendimiento cognitivo. Los resultados generales de este estudio publicado en un artículo semanal en 2013 amplían la noción de polimorfismos genéticos. Concluyen que el consumo regular de cannabis no es sustancialmente perjudicial en el rendimiento ejecutivo entre los usuarios de cannabis.

Se justifican los esfuerzos continuos para resolver el debate sobre la naturaleza de la asociación entre el cannabis y los posibles efectos adversos en las funciones ejecutivas. Cabe señalar que en el estudio realizado en 2013 no se aplicó puntajes de propensión para realizar el análisis de datos sino un modelo estadístico clásico, modelos ANCOVA.

Los resultados que arrojan los diferentes métodos del propensity score (Tabla 19) siguen el mismo patrón a excepción del método de estratificación por el propensity score que arroja dos patrones diferentes en contraste con los demás métodos. Los resultados de este estudio deben interpretarse a la luz de las posibles limitaciones metodológicas. Por ejemplo, una limitación del estudio es que no pudimos equilibrar una covariable medida, estado de fumar “smoke”. Los investigadores pueden dudar en aplicar técnicas de puntaje de propensión debido a la limitación principal de factores de confusión, falta de equilibrio suficiente para alguna de las covariables. Sin embargo, esta preocupación por los posibles factores de confusión no medidos también es común tanto en las técnicas de puntaje de propensión como en la regresión multivariante tradicional aplicada a los datos de observación. Por lo tanto, debido a que se aplica la misma limitación, no es una preocupación especial para este estudio en particular.

5.2. CONCLUSIONES

En este estudio se ha descrito el uso de cuatro diferentes métodos vanguardistas del propensity score para tratar datos observacionales. Y se ha implementado estas metodologías usando datos reales de un estudio realizado en 2013 sobre “COMT val158met and 5-HTTLPR Genetic Polymorphisms Moderate Executive Control in Cannabis Users”, con el objetivo de comparar la calidad de los resultados, verificar si hay cambios en los patrones y desenlaces finales.

Una de las fortalezas del propensity score es su capacidad para integrar y controlar muchas covariables, su uso es diverso, en inferencia causal se usa como un método para probar varios tecnicismos antes de ejecutar un modelo de resultado. Y una de las limitaciones principales del propensity score está basado en la especificación del modelo del propensity score y su posterior verificación de equilibrio en la distribución de las covariables entre los grupos de tratamiento, en el sentido, de simular las condiciones de identificabilidad, específicamente la intercambiabilidad, para que las inferencias causales sean válidas.

Los resultados obtenidos mediante matching, ponderación y el ajuste por el propensity score siguen el mismo patrón en comparación con los resultados del modelo ANCOVA. La estratificación por el propensity score, es el único que arrojó resultados que siguen patrones diferentes en dos de los índices de rendimiento de las tareas.

La evidencia de los análisis ajustados de puntaje de propensión, emparejamiento por el puntaje de propensión, estratificación por el puntaje de propensión y ponderación por el puntaje de propensión no respaldan el vínculo causal hipotético entre el uso de cannabis en los jóvenes y las funciones ejecutivas. Sobre la base de la evidencia de este estudio, los problemas de cannabis no parecen ser sustancialmente perjudicial en el rendimiento ejecutivo entre los usuarios de cannabis. En otras palabras, si podemos evitar que los jóvenes desarrollen problemas con el consumo de cannabis, notaríamos que hay poca reducción en la aparición o prevalencia de episodios de déficits neurocognitivos.

En este TFM hemos presentado métodos estadísticos que nos permiten acercarnos a la inferencia causal en estudios observacionales.

6. BIBLIOGRAFÍA

1. Austin, PC: A comparison of 12 algorithms for matching on the propensity score. *Statist. Med.* 2014, 33 1057–1069.
2. Austin PC: An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46:399–424, 2011.
3. Alex P., Delicado P., (2019), *Transparencias: Curso de Inferencia Avanzada*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.
4. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, and Stürmer T: Variable selection for propensity score models. *Am J Epidemiol.* 2006 June 15; 163(12): 1149–1156.
5. Cobo E, Videla S, Clotet B, Calle ML, Corchero C, Gómez G, Gruttola VD, Langohr K: *Investigación Clínica y Estadística*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.
6. Cochran WG, Cox GM: *Experimental Designs*. Wiley-Interscience, New York.
7. García AV, Fagundo AB, Cuenca A, Rodriguez J, Cuyás E, Langohr K, Llopis SS, Civit E, Farré M, Casanova JP and Torre R: COMT val158met and 5-HTTLPR Genetic Polymorphisms Moderate Executive Control in Cannabis Users. *Neuropsychopharmacology* (2013) 38, 1598–1606.
8. Hosmer DW, Lemeshow S: *Applied Logistic Regression*. Wiley-Interscience, New York.
9. Harder VS, Stuart EA, and Anthony JC: Adolescent Cannabis Problems and Young Adult Depression: Male-Female Stratified Propensity Score Analyses. *Am J Epidemiol* 2008;168:592–60.
10. Hernán MA, Robins JM: *Causal Inference*. Disponible en: https://cdn1.sph.harvard.edu/wpcontent/uploads/sites/1268/2020/02/ci_hernanrobins_21feb20.pdf.
11. Imbens GW: Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *The Review of Economics and Statistics*, February 2004, 86(1): 4-29.
12. Joffe MM and Rosenbaum PR: Invited Commentary: Propensity Scores. *Am J Epidemiol* Vol. 150, No. 4, 1999.
13. Kuss O, Blettner M, Börgermann J: Propensity Score: an Alternative Method of Analyzing Treatment Effects. *Dtsch Arztebl Int* 2016; 113: 597–603.

14. King G and Nielsen R: Why Propensity Scores Should Not Be Used for Matching. *Political Analysis* (2019).
15. Langohr K., (2019), *Transparencias: Curso de Epidemiologia*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.
16. Lunceford JK and Davidian M. Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study. *Statistics in Medicine*, 23, 2937– 2960.
17. Liu, L., & Ripley, D. (2014): Propensity Score Matching in a Study on Technology-Integrated Science Learning. *International Journal of Technology in Teaching and Learning*, 10 (2), 88-104.
18. Muriel A, Hernández D, Abaira V: Modelos estructurales marginales: una herramienta útil que proporciona evidencia a los estudios observacionales. *Nefrología Sup Ext* 2011;2(7):7-13.
19. Montero, L., (2014). *Apunts: Models Lineals Generalitzats*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.
20. Montero, L., (2004). *Apunts: Anàlisi de la Variança i Covariança versió 1.0*. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.
21. Martínez-González MA, Sánchez-Villegas A, Toledo-Atucha EA, Faulin-Fajardo J. *Bioestadística amigable* (2020). Elsevier España, S.L.U.
22. Rosenbaum PR and Rubin DB: The Bias Due to Incomplete Matching. *Biometrics*, Vol. 41, No. 1 (Mar., 1985), pp. 103-116.
23. Robins JM, Hernán MA and Brumback B: Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* September 2000, Vol. 11 No. 5.
24. Rosenbaum PR and Rubin DB: The central role of the propensity score in observational studies for causal effects. *Biometrika* (1983), 70, 1, pp. 41-5.
25. Rubin, D. B. (2006). *Matched sampling for causal effects*. New York, NY: Cambridge University Press.
26. Soler HM: Análisis de la Efectividad de un Tratamiento Mediante el Propensity Score. Disponible en: <https://upcommons.upc.edu/handle/2099.1/13214>.
27. Stampf S: Propensity score based data analysis. Disponible en: <https://www.semanticscholar.org/paper/Propensity-score-based-data-analysis-using-Stampf/1b5de084c812ae9d3ed847decde791ad7e96c202>.

28. Pattanayak CW, Rubin DB y Zell ER: Métodos de puntuación de propensión para crear una distribución equilibrada de las covariables en los estudios observacionales. Rev Esp Cardiol. 2011;64(10):897–903.
29. Pan W, Bai H: Propensity Score Analysis, Concepts and Issues. Disponible en: https://pdfs.semanticscholar.org/30e5/44dbc655d14b964638838ba410540ebfb3f5.pdf?_ga=2.95324565.860803549.1590781755-112044985.1570726939.
30. Pattanayak CW, Rubin DB y Zell ER: Métodos de puntuación de propensión para crear una distribución equilibrada de las covariables en los estudios observacionales. Rev Esp Cardiol. 2011;64(10):897–903.

- **Otros recursos:**

- <https://github.com/cran/nonrandom/tree/master/R#start-of-content>
- https://www.youtube.com/watch?v=AP_jA4mTaRs

7. APÉNDICE

```
## LOS PAQUETES NECESARIOS
```

```
## =====  
library(Hmisc)  
library(doBy)  
library(descr)  
library(car)  
library(MatchIt)  
library(lme4)  
source("Rfunction_summby.R")
```

7.1. APÉNDICE 1

```
##=====
```

Análisis descriptivo (con prueba de hipótesis opcional) de un marco de datos para cada categoría de una variable categórica.

```
##=====
```

```
## INPUT:
```

1. daten: marco de datos
2. catvar: variable categórica de daten
3. namcv: nombre de catvar para output de tablas de resumen
4. Ftes: F-test para variables numéricas (Default: FALSE)
5. Chi: Chisquare-test para variables categóricas (Default: FALSE)
5. titl: Título para el output (Default: NULL)
7. vert: Porcentajes en tablas de contingencia para catvar (FALSE) o para variable categorica (TRUE; default value)
8. repo: Incluir etiquetas variables en el output (Default: TRUE)

```
##=====
```

Bucle: Función summby: Proporcionado por, Klaus Langohr

```
## =====
```

```
## Función summby()
```

```
summby <- function(daten, catvar, namcv, Ftes = FALSE, Chi = FALSE, titl  
= NULL,
```

```
vert = TRUE, repo = TRUE) {
```

```
# Auxilliary function to underline text
```

```
underline <- function(txt) {
```

```
  cat("\n", txt, fill = TRUE, sep = "")
```

```
  cat(rep("=", nchar(txt)), fill = TRUE, sep = "")
```

```
}
```

```
# Auxilliary function sumfun for descriptive statistics
```

```
sumfun <- function(x, ...) {
```

```

      c(N = sum(complete.cases(x)), Missings = sum(is.na(x)), m = mean(x,
...),
      sd = sd(x, ...), Med = median(x, ...), Min = min(x, ...), Max =
max(x, ...))
    }

    if (!is.null(titl)) {
      underline(titl)
    }

    catvindex <- which(colnames(daten) == catvar)
    colnames(daten)[catvindex] <- "catv"

    nums <- which(sapply(daten, is.numeric))
    kats <- which(sapply(daten, function(x) is.factor(x) |
is.character(x)))
    kats <- kats[-which(names(kats) == names(daten)[catvindex])]
    for (i in nums) {
      daten$vari <- daten[, i]
      if (repo) {
        underline(label(daten[, i]))
      }
      sumby <- summaryBy(vari ~ catv, daten, FUN = sumfun, na.rm = TRUE)
      names(sumby) <- c(namcv, "N", "Missings", "Mean", "SD", "Median",
"Min.",
                        "Max.")
      print(sumby, digits = 3)
      if (Ftes) {
        efi <- oneway.test(vari ~ catv, daten, var.equal = TRUE)
        cat("\nF-test: F = ", round(efi$stat, 3), "; df = ", efi$para[1], "
& ",
          efi$para[2], "; p = ", round(efi$p.val, 3), ".\n", sep = "")
      }
      daten$vari <- NULL
    }
    rm(i)
    if (vert) {
      for (i in kats) {
        daten$vari <- daten[, i]
        if (repo) {
          underline(label(daten[, i]))
        }
        print(CrossTable(daten$vari, daten$catv, prop.r = FALSE, prop.t =
FALSE,
                        prop.chisq = FALSE, cell.layout = FALSE, format =
"SPSS",
                        row.labels = FALSE, dnn = c("", namcv)))
        if (Chi) {
          chis <- with(daten, suppressWarnings(chisq.test(vari, catv)))

```

```

        cat("Chi-squared test: Chi-squared = ", round(chis$stat, 3), ";
df = ",
        chis$para, "; p = ", round(chis$p.val, 3), ".\n", sep = "")
    }
    daten$vari <- NULL
  }
} else {
  for (i in kats) {
    daten$vari <- daten[, i]
    if (repo) {
      underline(label(daten[, i]))
    }
    print(CrossTable(daten$vari, daten$catv, prop.c = FALSE, prop.t =
FALSE,
                    prop.chisq = FALSE, cell.layout = FALSE, format
= "SPSS",
                    row.labels = FALSE, dnn = c("", namcv)))
    if (Chi) {
      chis <- with(daten, suppressWarnings(chisq.test(vari, catv)))
      cat("Chi-squared test: Chi-squared = ", round(chis$stat, 3), ";
df = ",
      chis$para, "; p = ", round(chis$p.val, 3), ".\n", sep = "")
    }
    daten$vari <- NULL
  }
}
}

##=====

```

```

##=====
## Modelos ANCOVA: Usado por Verdejo et al. (2013),
## Códigos proporcionado por: Klaus Langohr
##=====
## Bucle para hacer análisis descriptivo y los modelos anova
## -----
## Las variables de interés

vars <- c("rvptprob", "rvpa", "rvpb", "nback2dif", "nback4dif",
"nback6dif",
        "iedterradj", "iediderrors", "iedederrors", "otsmchoice",
"otsmlatfirst",
        "igtbloque1", "igtbloque2", "igtbloque3", "igtbloque4",
"igtbloque5",
        "igtttotal", "zscore")

for (v in seq_along(vars)) {
  cannabi$var <- cannabi[, vars[v]]

  ## Análisis descriptivo
  sumby(cannabi[, c("group", "var")], "group", "Group")

  ## Modelo ANCOVA
  ancova <- lm(var ~ group + gender + age + civoc + smoke, cannabi)

  ## Estimación de parámetros del modelo ANCOVA
  cat("\n")
  cat("Parameter estimates", fill = TRUE)
  cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
  print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

  ## Tabla ANOVA del modelo ANCOVA
  cat("\n")
  print(round>Anova(ancova), 3), signif.stars = FALSE)
  cannabi$var <- NULL
  rm(ancova)
}
rm(v, vars)
##=====

```

7.2. APÉNDICE 2

```
## Más funciones para realizar análisis no aleatorio (nonrandom)
## De: Susanne Stampf (2014)
##=====
## Función para la estimación de la densidad del PS estimado
##=====

pscore <- function(formula,
                    data,
                    family=binomial,
                    na.action=na.exclude,
                    name.pscore="pscore",
                    ...
)
{
  ## Check data argument
  if (missing(data)){
    stop("Argument 'data' is missing.")
  }else{
    if(!inherits(data,"data.frame"))
      stop("Argument 'data' is not of class 'data.frame'.")
  }

  ## Check formula argument
  if (missing(formula)){
    stop("Argument 'formula' is missing.")
  }else{
    if(!inherits(formula,"formula"))
      stop("Argument 'formula' is not of class 'formula'.")
  }

  ## Check arguments if necessary
  if(any(names(data)==name.pscore))
    stop(paste("Variable 'name.pscore'=",name.pscore," already exists in
data.", sep=""))

  ## Extract treatment ( == response here)
  name.treat <- names(model.frame(formula,data))[1]
  treat      <- data[,name.treat]

  ## Fit PS model
  ps.model <- glm(formula, family, data, na.action=na.action, ...)

  ## Predict PS values
  data[,name.pscore] <- predict(ps.model, type="response")

  ## Define output:
  pscore <- data[,name.pscore]
```

```

output <- list(data      = data,
               pscore    = pscore,
               name.pscore = name.pscore,
               formula.pscore = formula,
               model.pscore = ps.model,
               treat      = treat,
               name.treat = name.treat)

class(output) <- c("pscore")

return(output)

}

##=====
## Función para la estimación del propensity score
##=====

plot.pscore <- function(x,
                        par.dens=NULL,
                        par.1=NULL,
                        par.0=NULL,
                        with.legend=FALSE,
                        legend.cex=0.9,
                        legend.label=NULL,
                        main=NULL,
                        ylim=NULL,
                        xlim=NULL,
                        ...)
{
  object <- x

  if ( any(class(object) == "pscore") ){

    require( "colorspace", character.only=TRUE )

    if (is.null(ylim))
      ylim <- c(0, ceiling(max(density(object$pscore[object$treat==1])$y,
density(object$pscore[object$treat==0])$y)))
    if (is.null(xlim))
      xlim <- c(floor(min(density(object$pscore[object$treat==1])$x,
density(object$pscore[object$treat==0])$x)),
ceiling(max(density(object$pscore[object$treat==1])$x,
density(object$pscore[object$treat==0])$x)))

    plot(density(object$pscore),
         col="white",
         main=main,
         ylim=ylim,

```



```

        xlim=xlim,
        ...)

x1 <- do.call("density",
              c(list(x=object$pscore[object$treat==1]),
                par.dens))
x0 <- do.call("density",
              c(list(x=object$pscore[object$treat==0]),
                par.dens))

if (is.null(par.1) & is.null(par.0))

  par.0 <- list(lty=2)

do.call("lines", c(list(x=x1), par.1))
do.call("lines", c(list(x=x0), par.0))

if (with.legend){

  if (is.null(legend.label)){
    legend.label <- c("Control", "Cannabis")}

  legend(x=xlim[2]/4,
        y=ylim[2],
        cex=legend.cex,
        lty=c(1:2),
        legend.label,
        ncol=2)
  }
}else{
  stop("Function 'pscore' must be at least used before.")
}
}

###=====
## Función para realizar el MATCHING por el propensity score
##=====
ps.match <- function(object,
                     object.control      = NULL,
                     matched.by          = NULL,
                     control.matched.by  = matched.by,
                     who.treated         = 1,
                     treat               = NULL,
                     name.match.index    = "match.index",
                     ratio               = 1,
                     caliper             = "logit",
                     x                   = 0.2,
                     givenTmatchingC    = TRUE,
                     bestmatch.first     = TRUE,

```

```

        setseed          = FALSE,
        combine.output    = TRUE)
{
  if (missing(object))
    stop("Argument 'object' is not given.")

  if (any(class(object)=="pscore") |
      any(class(object)=="data.frame"))

    UseMethod("ps.match")
  else
    stop("Class of argument 'object' will not supported.")
}
##=====

###=====
## Función para realizar la verificación de equilibrio en la distribución
## de las covariables mediante gráficos
##=====
dist.plot<- function(object,
                      sel          = NULL,
                      treat        = NULL,
                      stratum.index = NULL,
                      match.index  = NULL,
                      plot.type    = 1,
                      compare      = FALSE,
                      cat.levels   = 2,
                      plot.levels  = 5,
                      label.match  = c("original data","matched sample"),
                      label.stratum = c("Stratum","Original"),
                      with.legend  = TRUE,
                      legend.title = NULL,
                      legend.cex   = 1.5,
                      myoma        = c(3,2,2,2),
                      mymar        = c(5,4,1,2),
                      width        = 0.9,
                      xlim         = NULL,
                      ylim         = NULL,
                      col          = NULL,
                      las          = 1,
                      font.main    = 2,
                      font         = 1,
                      main         = NULL,
                      main.cex     = 1.2,
                      sub.cex      = 0.9,
                      bar.cex      = 0.8,
                      ...
)
{

```

```

if(missing(object))

  stop("Argument 'object' is needed.")

else

  if (class(object)[1]=="stratified.data.frame" |
      class(object)[1]=="stratified.pscore" |
      class(object)[1]=="matched.data.frame" |
      class(object)[1]=="matched.data.frames" |
      class(object)[1]=="matched.pscore" |
      class(object)[1]=="data.frame" |
      class(object)[1]=="pscore")

    UseMethod("dist.plot")

  else

    stop("Class of argument 'object' will not supported.")

}
##=====

###=====
## Función para LA VERIFICACION DE EQUILIBRIO MEDIANTE PRUEBAS
## ESTADÍSTICAS
##=====
ps.balance <- function(object,
                        sel          = NULL,
                        treat        = NULL,
                        stratum.index = NULL,
                        match.index  = NULL,
                        method       = "classical",
                        cat.levels   = 2,
                        alpha        = 5,
                        equal        = TRUE)
{
  if (missing(object))
    stop("Argument 'object' is not given.")

  if (any(class(object)=="stratified.data.frame") |
      any(class(object)=="stratified.pscore") |
      any(class(object)=="matched.data.frame") |
      any(class(object)=="matched.data.frames") |
      any(class(object)=="matched.pscore") |
      any(class(object)=="data.frame"))

```

```

    UseMethod("ps.balance")
  else
    stop("Class of argument 'object' will not supported.")
}
##=====

###=====
## Función para la estratificación por el propensity score
##=====
ps.makestrata <- function(object,
                          breaks          = NULL,
                          name.stratum.index = "stratum.index",
                          stratified.by    = NULL,
                          ...
)
{
  if (missing(object))
    stop("Argument 'object' is not given.")

  if (any(class(object)=="pscore") |
      any(class(object)=="data.frame"))

    UseMethod("ps.makestrata")
  else
    stop("Class of argument 'object' will not supported.")
}
##=====

###=====
## Función para realizar la gráfica de diferencias estandarizadas en la
## verificación de equilibrio
##=====
plot.stdF <- function(x,
                      sel          = NULL,
                      plot.alpha   = TRUE,
                      mymar        = c(5,8,4,2),
                      pch.p        = c(1,5),
                      col.p        = c("blue", "red"),
                      colorspace   = NULL,
                      cex.p        = 1.25,
                      line.stdF    = 1,
                      line.alpha   = 4,
                      with.legend  = TRUE,
                      legend.label = c("before matching", "after
matching"),
                      legend.cex   = 1,

```

```

        legend.xy      = NULL,
        ...)
{
  object <- x

  ## #####
  ## check colorspace
  if (!is.null(colorspace)){
    suppressWarnings(require( "colorspace", character.only=TRUE ))
    message("Argument 'col.p' is ignored unless it has numeric values.")

    if (colorspace==TRUE){
      if (any(apply(as.data.frame(col.p),1,is.numeric))==FALSE){
        col.p <- sample(rainbow_hcl(20),2)
        message("Plot colors are randomly chosen since 'col.p' are
strings.")
      }else{
        col.p <- rainbow_hcl(20)[col.p]
      }
    }else{

      if (colorspace==FALSE){
        if (any(apply(as.data.frame(col.p),1,is.numeric))==FALSE){
          col.p <- sample(grey.colors(20),2)
          message("Plot colors are randomly chosen since 'col.p' are
strings.")
        }else{
          col.p <- grey.colors(20)[col.p]
        }
      }else{
        warning("Argument 'colorspace' is ignored since it is not
logical.")
      }
    }
  }

  ## #####
  ## check object
  if(missing(object)){
    stop("Argument 'object' is needed.")
  }else{
    if (any(substring(class(object)[1],1,7) == "bal.str")){
      stop("Matching is not done before.")
    }else{
      if (is.null(object$bal.test$Stand.diff)){
        stop("No standardized differences are available.")
      }else{

```

```

        if (dim(object$bal.test$Stand.diff)[1] != 2)
          stop("Matching is not done before.")
      }
    }
  }

## #####
## find sel, only names are of interest
bal.var <- colnames(object$bal.test$Stand.diff)

if(is.null(sel)){
  name.sel <- bal.var
}else{
  sel <- find.sel(data=object$data,
                  sel=sel)
  name.sel <- unique(names(sel))

  if (length(intersect(name.sel, bal.var)) == 0){
    stop("No standardized differences available for all selected
variables.")
  }else{
    if (length(name.sel) > length(intersect(name.sel, bal.var))){
      cat("There are no standardized differences for selected
variable(s):\n")
      print(setdiff(name.sel, bal.var))
      name.sel <- intersect(name.sel, bal.var)
    }
  }
}

## ####
## plot
val.b <- object$bal.test$Stand.diff[1,][name.sel]
val.a <- object$bal.test$Stand.diff[2,][name.sel]
stdf <- object$bal.test$Stand.diff[,name.sel]

## recode stdf for variables with Inf
if (any(stdf==Inf)) stdf[stdf==Inf] <- NaN

par(mar=mymar)
plot(val.b,
     2:(length(name.sel)+1),
     pch=pch.p[1],
     col=col.p[1],
     cex=cex.p,
     axes=F,
     ylim=c(1, length(name.sel)+1.5),
     xlim=c(min(floor(min(round(stdf, 2), na.rm=TRUE)),0),

```

```

        ceiling(max(round(stdf, 2), na.rm=TRUE))),
xlab="",
ylab="",
...)

points(val.a,
       2:(length(name.sel)+1),
       pch=pch.p[2],
       col=col.p[2],
       cex=cex.p)
axis(1,
     ...)
axis(2,
     at=2:(length(name.sel)+1),
     labels=name.sel,
     ...)
box()

apply(as.data.frame(cbind(val.a, val.b, c(2:(length(name.sel)+1)))), 1,
      function(x)
        lines(c(x[1],x[2]), c(x[3],x[3]), lty=line.stdf))

if (plot.alpha)
  abline(v=object$bal.test$alpha, lty=line.alpha)

if (with.legend){
  if (is.null(legend.xy)){
    legend(x      = ceiling(max(round(stdf, 2), na.rm=TRUE))/2,
          y      = 1.5,
          legend  = legend.label,
          horiz  = TRUE,
          pch     = pch.p,
          col     = col.p,
          pt.cex  = cex.p,
          cex     = legend.cex)
  }else{
    if ( !is.numeric(legend.xy) ){
      stop("Argument 'legend.xy' must be numeric.")
    }else{
      if (length(legend.xy) != 2 ){
        stop("Argument 'legend.xy' must be a numeric vector of length
2.")
      }else{
        legend(x      = legend.xy[1],
              y      = legend.xy[2],
              legend  = legend.label,
              horiz  = TRUE,
              pch     = pch.p,
              col     = col.p,
              pt.cex  = cex.p,

```

```

                                cex      = legend.cex)
                                }
                                }
                                }
                                }
                                }

}

##=====

```

7.3. APÉNDICE 3

```

###=====
                                ## APLICACIÓN DEL PROPENSITY SCORE:
##=====
## CARGAR LOS DATOS
## =====
## Cargamos los datos
## ++++++
load("Verdejo2013.RData")
Label(cannabi)
summary(cannabi)
dim(cannabi)
## SEGUNDO: Seleccionamos las variables de interés
datos<-cannabi[,c(1:5,13:31)]
cannabi6<-datos[,c(1:5,6,7,9,10,13,23,24)]
## TERCERO: Eliminar missing
datos<- na.omit(cannabi6)
#CUARTO:Se asigna valores binarios a los niveles de group
datos$group<- as.numeric(datos$group) - 1
datos$gender <- as.numeric(datos$gender) - 1
datos$smoke <- as.numeric(datos$smoke) - 1
##=====

```



```

####=====
## ESTIMACIÓN DEL PROPENSITY SCORE
## Las covariables de interés para la estimación de PS son: GENDER, AGE,
## CIVOC, SMOKE Estas variables se ajustan en modelo de regresión
## logística con variable de tratamiento GROUP.
####=====
## Estimación del propensity score (PS)

cannabi.ps <- pscore(data = datos, formula = group ~ gender + age +
civoc + smoke)

PS<-cannabi.ps$pscore
hist(PS, main = "Histograma del propensity score estimado")
print.pscore(cannabi.ps)

# Gráfico 1: Estimación de la densidad del PS estimado
plot.pscore(cannabi.ps, par.1=list(lty=1,lwd=2),
            par.0=list(lty=2,lwd=2),
            with.legend=TRUE,
            main="Densidad del propensity score",
            ylim=c(0,2.5),
            xlim=c(0,1),
            xlab="PS")

plot(cannabi.ps, with.legend=T, xlim=c(0,1))

##histograma cannabis y control

data<-cannabi.ps$data
PS<-data$pscore
logitPS<-(log(PS/(1-PS)))
my_data<-cbind(data,logitPS) ## datos con pscore y logit del pscore

par(mfrow=c(1,1))
hist(my_data$pscore[my_data$group==1],col="blue", border="black")
hist(my_data$pscore[my_data$group==0] ,col="red", border="black")
hist(my_data$pscore,col="yellow", border="black")

#summary(my_data$pscore[my_data$group==1])
#sd(my_data$pscore[my_data$group==1])
summary(my_data$pscore[my_data$group==0])
sd(my_data$pscore[my_data$group==0])

par(mfrow=c(1,1))
hist(my_data$logitPS[data$group==1],col="blue", border="black")
hist(my_data$logitPS[data$group==0] ,col="red", border="black")
hist(my_data$logitPS,col="yellow", border="black")

####=====

```

```
###+++++
```

MÉTODOS BASADOS EN PS

```
###=====
```

```
## El emparejamiento y la estratificación por PS se usa para eliminar los  
## desequilibrios en las distribuciones covariables, creando una  
## situación de datos equilibrados como en los ensayos controlados  
## aleatorios.
```

```
###=====+++++
```

PROPENSITY SCORE MATCHING

```
###=====+++++
```

```
## Es el método más utilizado para equilibrar la distribución de las  
## covariables. Los individuos de cannabis se comparan con los individuos  
## de control en una proporción de 1:1 (emparejamiento 1:1) utilizando el  
## tamaño de caliper predeterminado. La variable de emparejamiento se  
## establece en "ps".
```

```
##=====
```

EMPAREJAMIENTO DE CALIPER

```
###+++++
```

```
# Emparejamiento 1 (51x51)
```

```
cannabi.m1 <- ps.match(object = cannabi.ps, ratio = 1, x = 0.2,  
                      caliper = "logit", matched.by =  
"pscore", set.seed(38902))
```

```
summary(cannabi.m1)
```

```
# Emparejamiento 2 (40x40)
```

```
#cannabi.m1 <- ps.match(object = cannabi.ps, ratio = 1, x = 0.2, caliper  
=0.2,  
                      #matched.by = "pscore", set.seed(38902))
```

```
#summary(cannabi.m1)
```

```
# Emparejamiento 3 (56x56)
```

```
#cannabi.m1 <- ps.match(object = cannabi.ps, ratio = 1, x = 0.2, caliper  
=0.4,  
                      #matched.by = "pscore", set.seed(38902))
```

```
#summary(cannabi.m1)
```

```
# Emparejamiento 3 (34x34)
```

```
#cannabi.m1 <- ps.match(object = cannabi.ps, ratio = 1, x = 0.2, caliper  
=0.09,  
                      #matched.by = "pscore", set.seed(38902))
```

```
#summary(cannabi.m1)
```

```
#m1<-cannabi.m1$data.matched
```

```

##=====
##                                EMPAREJAMIENTO DE CALIPER=0.5
##+++++
+++++
# Emparejamiento 2
#cannabi.m2 <- ps.match(object = cannabi.ps, ratio = 2, caliper =0.5,
                        #givenTmatchingC = FALSE, setseed = 2000)

#summary(cannabi.m2)

## VERIFICACIÓN DE EQUILIBRIO DE LAS COVARIABLES
##+++++
+++++

## Gráfico de histogramas para variables continuas
dplot1<- dist.plot(object = cannabi.m1, sel =c("age"),
                  compare = TRUE, plot.type=2,with.legend=T)

dplot2<- dist.plot(object = cannabi.m1, sel =c("civoc"),
                  compare = TRUE, plot.type=2,with.legend=T)

## Gráfico de barras apiladas para variables categóricas

windows(width=10, height = 5)
dplot3<- dist.plot(object = cannabi.m1, sel =c("smoke"),
                  compare = TRUE, label.match=c("original data","matched
sample"),mymar=c(2,2,1,2))

dplot4<- dist.plot(object = cannabi.m1, sel =c("gender"),
                  compare = TRUE, label.match=c("original data","matched
sample"),mymar=c(2,2,1,2))

windows(width=15, height = 10)
dplot3<- dist.plot(object = cannabi.m1, sel =c("gender"),
                  compare = TRUE, label.match=c("original data","matched sample"),
                  mymar=c(2,2,2,2),legend.title="Cannabis vs Control",
                  main.cex=5,legend.cex = 1.2,sub.cex=1.2, bar.cex = 1.2)
## verificacion mediante proporciones para variables categoricas
##+++++
## para la variable categórica gender
dplot3$var.cat
dplot3$frequency

## para la variable categórica gender
dplot4$var.cat
dplot4$frequency

```

```

# verificacion mediante proporciones para variables continuas
## para la variable continua age
#####
dplot1$var.noncat
dplot1$breaks.noncat ## puntos de corte para clase de histograma
dplot1$x.s.noncat    ## frecuencias para el grupo 0 en la data original y
                      ## en la muestra emparejada

dplot1$y.s.noncat    ## frecuencia para el grupo 1 en la data original
                      ## en la muestra emparejada

## para la variable continua civoc
dplot2$var.noncat
dplot2$breaks.noncat ## puntos de corte para clase de histograma
dplot2$x.s.noncat    ## frecuencias para el grupo 0 en la data original
y
                      ## en la muestra emparejada

dplot2$y.s.noncat    ## frecuencia para el grupo 1 en la data original
                      ## en la muestra emparejada

## Pruebas estadísticas y diferencias estandarizadas
#####
#cannabi.m1 <- ps.match(object = cannabi.ps, ratio = 1, x = 0.2,
                        #caliper = "logit", matched.by =
"pscore", set.seed(38902))

#summary(cannabi.m1)

cannabi.m1.bal <- ps.balance(object = cannabi.m1, sel = c(3,4,5,11),
                           cat.levels = 2, method = "classical",
alpha = 5)

summary(cannabi.m1.bal)

## diferencias estandarizadas
#####
cannabi.m1.bal_est <- ps.balance(object = cannabi.m1, sel =
c("gender", "age", "civoc", "smoke"),
                              method = "stand.diff", alpha = 10)
cannabi.m1.bal_est

distplot_est <- plot.stdf(x=cannabi.m1.bal_est,
sel=c("gender", "age", "civoc", "smoke"), main="cannabi: Matching by
PS", pch.p = c(20,10))

```

```

##+++++
## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA (Emparejamiento)
## =====
###+++++
# Muestra emparejada
m1<-cannabi.m1$data.matched

## PASO 1: LAS VARIABLES DE INTERÉS (Variables dependientes)

vars <- c("rvptprob","nback2dif","iedterradj","otsmlatfirst","igtttotal",
"zscore")

## PASO 2: BUCLE PARA HACER ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
for (v in seq_along(vars)) {
m1$var <- m1[, vars[v]]

## ANÁLISIS DESCRIPTIVO
sumby(m1[, c("group", "var")], "group", "Group")

## MODELO ANCOVA (ajuste por emparejamiento)
ancova<- lm(var ~ group + gender + age + civoc + smoke, m1)

## ESTIMACION DE PARÁMETROS DEL MODELO ANCOVA
cat("\n")
cat("Parameter estimates", fill = TRUE)
cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

## TABLA ANOVA DEL MODELO ANCOVA
cat("\n")
print(round(Anova(ancova), 3), signif.stars = FALSE)
m1$var <- NULL
rm(ancova)
}
rm(v, vars)

```

```

####=====
### ESTRATIFICACIÓN POR EL PROPENSITY SCORE
####=====
## Estratificación por PS
## Agrupa a individuos que tienen valores similares o iguales para la
## variable de estratificación (PS).

#=====
## (1) Estratificación (factorización)
#++++++
cannabi.str74<- ps.makestrata.pscore(object =cannabi.ps)
#summary.stratified.pscore(cannabi.str74)

## (2) Estratificación (especificación del número de
estratos)
#++++++
cannabi.str.b3 <- ps.makestrata(object = cannabi.ps, breaks = 3)
summary.stratified.pscore(cannabi.str.b3)

## (3) Estratificación (definición de límites de estratos específicos)
#++++++
+++++

# ESTRATIFICACIÓN 1: QUINTILES DEL PROPENSITY SCORE (PS)
cannabi.str5 <- ps.makestrata(object = cannabi.ps, breaks =
quantile(cannabi.ps$pscore,
seq(0,1,0.2)),name.stratum.index =
"stratum")

summary.stratified.pscore(cannabi.str5)

# ESTRATIFICACIÓN 2: DECILES DEL PROPENSITY SCORE (PS)
#cannabi.str10 <- ps.makestrata(object = cannabi.ps, breaks =
quantile(cannabi.ps$pscore,
#seq(0,1,0.1)),name.stratum.index =
"stratum")

#summary.stratified.pscore(cannabi.str10)

#=====
## VERIFICACION DE EQUILIBRIO MEDIANTE GRÁFICO DE BARRAS
##+++++
+++++
## Gráfico: Verificacion de equilibrio (estratificacion por quintiles)
#windows(width=10, height = 5)
#cannabi.dplot1 <- dist.plot(object = cannabi.str5, sel = "age",
#label.stratum =
c("Str.,""Original"),plot.type=1, compare=TRUE)

```

```

## Gráfico: Verificación de equilibrio (estratificación por deciles)
#windows(width=10, height = 5)
#cannabi.dplot2 <- dist.plot(object = cannabi.str10, sel = "age",
                             #label.stratum =
c("Str.", "Original"), plot.type=1, compare=TRUE)

windows(width=10, height = 5)
pride.plot3 <- dist.plot(object = cannabi.str5, sel = c("gender"),
plot.type = 1,
                             compare=TRUE, label.stratum =
c("Str.", "Original"), mymar=c(1,1,1.5,2))

pride.plot3$frequency

windows(width=10, height = 5)
pride.plot4 <- dist.plot(object = cannabi.str5, sel = c("smoke"),
plot.type = 1,
                             compare=TRUE, label.stratum =
c("Str.", "Original"), mymar=c(1,1,1.5,2))

pride.plot4$frequency

#windows(width=10, height = 5)
#cannabi.dplot5 <- dist.plot(object = cannabi.str5, sel = "civoc",
                             #label.stratum = c("Str.", "Original"), plot.type=1,
compare=TRUE)

## verificación de equilibrio para variables categóricas (frecuencias)
pride.plot3$frequency

pride.plot4$frequency

## verificación de equilibrio para variables continuas (medias)
dplot11 <- dist.plot(object = cannabi.str5, sel = c("age"),
                     compare = TRUE, plot.type=2, with.legend=T)

dplot55 <- dist.plot(object = cannabi.str5, sel = c("civoc"),
                     compare = TRUE, plot.type=2, with.legend=T)

## frecuencia para EDAD
dplot11$var.noncat
dplot11$breaks.noncat ## puntos de corte para clase de histograma
dplot11$x.s.noncat     ## frecuencia para el GRUPO 0 en la data original y

```

```

## en la muestra estratificada
dplot11$y.s.noncat ## frecuencia para el GRUPO 1 en la data original
## en la muestra estratificada

## frecuencia para EDAD
dplot11$var.noncat
dplot11$breaks.noncat ## puntos de corte para clase de histograma
dplot11$x.s.noncat ## frecuencia para el GRUPO 0 en la data original y
## en la muestra estratificada

dplot11$y.s.noncat ## frecuencia para el GRUPO 1 en la data original
## en la muestra estratificada

###=====
====
## VERIFICACIÓN DE EQUILIBRIO MEDIANTE LAS MEDIAS
##+++++
++++
## Verificación de equilibrio para variables continuas (medias)
cannabi.dplot1$var.noncat
cannabi.dplot1$mean
cannabi.dplot5$var.noncat
cannabi.dplot5$mean

###=====
## VERIFICACIÓN DE EQUILIBRIO MEDIANTE PRUEBAS ESTADÍSTICAS
#+++++

## Verificación de equilibrio (pruebas estadísticas) para estratificación
cannabi.str5.bal <- ps.balance(object = cannabi.str5, sel = c(3,4,5,11),
                             cat.levels = 2, method = "classical",
alpha = 5)

summary(cannabi.str5.bal)

## diferencias estandarizadas para estratificación
cannabi.str5.bal_est <- ps.balance(object = cannabi.str5, sel =
c("gender","age","civoc","smoke"),
                             method = "stand.diff", alpha = 10)
cannabi.str5.bal_est

distplot_est<-plot.std(x=cannabi.str5.bal_est,
sel=c("gender","age","civoc","smoke"),main="cannabi: Matching by
PS",pch.p = c(20,10))

```



```

##+++++
## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
## =====
###=====
### MÉTODO 2: ESTRATIFICACIÓN POR EL PROPENSITY SCORE
## =====
###+++++

# Muestra estratificada
str5<-cannabi.str5$data
str5$stratum<- as.numeric(str5$stratum)
str(str5)

## PASO 1: LAS VARIABLES DE INTERÉS (Variables dependientes)

vars <- c("rvptprob","nback2dif","iedterradj","otsmlatfirst","igttotal",
"zscore")

## PASO 2: BUCLE PARA HACER ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
for (v in seq_along(vars)) {
  str5$var <- str5[, vars[v]]

  ## ANÁLISIS DESCRIPTIVO
  summbly(str5[, c("group", "var")], "group", "Group")

  ## MODELO ANCOVA (ajuste por emparejamiento)
  ancova<- lm(var ~ group + gender + age + civoc + smoke, weights =
stratum, str5)

  ## ESTIMACION DE PARÁMETROS DEL MODELO ANCOVA
  cat("\n")
  cat("Parameter estimates", fill = TRUE)
  cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
  print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

  ## TABLA ANOVA DEL MODELO ANCOVA
  cat("\n")
  print(round(Anova(ancova), 3), signif.stars = FALSE)
  str5$var <- NULL
  rm(ancova)
}
rm(v, vars)

##+++++

```

```

## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
## =====
###=====
### MÉTODO 3: PONDERACIÓN POR EL PROPENSITY SCORE
## =====
###+++++

## PASO 1: ASIGNACIÓN DE LOS PESOS (1/PS y 1/1-PS)
datos$gweight<-ifelse(datos$group== 1,1/datos$PS,1/(1-datos$PS))

## LAS VARIABLES DE INTERÉS(variables dependientes)
vars <- c("rvptprob","nback2dif","iedterradj","otsmlatfirst","igtttotal",
"zscore")

## PASO 2: BUCLE PARA HACER ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
for (v in seq_along(vars)) {
datos$var <- datos[, vars[v]]

## ANÁLISIS DESCRIPTIVO
sumby(datos[, c("group", "var")], "group", "Grupo")

## MODELO ANCOVA (ajuste por ponderación)
ancova<-lm(var~group + gender + age + civoc + smoke, data=datos,weights =
gweight)

## ESTIMACION DE PARÁMETROS DEL MODELO ANCOVA
cat("\n")
cat("Parameter estimates", fill = TRUE)
cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

## TABLA ANOVA DEL MODELO ANCOVA
cat("\n")
print(round(Anova(ancova), 3), signif.stars = FALSE)
datos$var <- NULL
rm(ancova)
}
rm(v, vars)
###+++++
###=====

```

```

##+++++
## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
## =====
###=====
### MÉTODO 4: PROPENSITY SCORE COMO VARIABLE DE AJUSTE
###=====
##+++++
## LAS VARIABLES DE INTERÉS (variables dependientes)
vars <- c("rvptprob","nback2dif","iedterradj","otsmlatfirst","igttotal",
"zscore")

## PASO 2: BUCLE PARA HACER ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
for (v in seq_along(vars)) {
datos$var <- datos[, vars[v]]

## ANÁLISIS DESCRIPTIVO
sumby(datos[, c("group", "var")], "group", "Grupo")

## MODELO ANCOVA (Ajuste por el propensity score)
ancova<- lm(var ~ group + PS, datos)

## ESTIMACION DE PARÁMETROS DEL MODELO ANCOVA
cat("\n")
cat("Parameter estimates", fill = TRUE)
cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

## TABLA ANOVA DEL MODELO ANCOVA
cat("\n")
print(round(Anova(ancova), 3), signif.stars = FALSE)
datos$var <- NULL
rm(ancova)
}
rm(v, vars)
##+++++
##=====

```

```

##+++++
## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
## =====
## ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA (CON DATOS ORIGINALES)
## =====
##+++++
  datos  ## datos de estudio
## PASO 1: LAS VARIABLES DE INTERÉS (Variables dependientes)

vars <- c("rvptprob","nback2dif","iedterradj","otsmlatfirst","igtttotal",
"zscore")

## PASO 2: BUCLE PARA HACER ANÁLISIS DESCRIPTIVO Y LOS MODELOS ANOVA
for (v in seq_along(vars)) {
  datos$var <- datos[, vars[v]]

  ## ANÁLISIS DESCRIPTIVO
  sumby(datos[, c("group", "var")], "group", "Group")

  ## MODELO ANCOVA (ajuste por emparejamiento)
  ancova<- lm(var ~ group + gender + age + civoc + smoke, datos)

  ## ESTIMACION DE PARÁMETROS DEL MODELO ANCOVA
##+++++
  cat("\n")
  cat("Parameter estimates", fill = TRUE)
  cat(rep("-", nchar("Parameter estimates")), sep = "", fill = TRUE)
  print(round(summary(ancova)$coefficients, 3), signif.stars = FALSE)

  ## TABLA ANOVA DEL MODELO ANCOVA
  cat("\n")
  print(round(Anova(ancova), 3), signif.stars = FALSE)
  datos$var <- NULL
  rm(ancova)
}
rm(v, vars)
##+++++
##=====

```

7.4. APÉNDICE 4

```
##+++++
##AXULIAR: Implementación mediante el paquete matchit en R
###=====
##=====
      ## ESTIMACIÓN DEL PROPENSITY SCORE (PS)
      ##=====
####=====

# PASO 1: ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO DE REGRESIÓN LOGÍSTICA
+++++
+++++
param<-glm(data=datos, group~gender + age + civoc + smoke, family =
"binomial")
param

#+++++
      # PASO 2: ESTIMACIÓN DEL PROPENSITY SCORE (PS)
#+++++
datos$PS<- predict(param, type="response")

#+++++
prob_tra<-data.frame(pscore=predict(param, type="response"),
group=param$model$group)
prob_tra[1:5,]

## PROPENSITY SCORE MATCHING
## =====

# PASO 1: EMPAREJAMIENTO DE CALIPER
##
+++++
+++++
# muestra emparejada 51x51
matching<-matchit(data=datos,group~gender + age + civoc +
smoke,caliper=1.12,
                  method ="nearest",ratio=1,distance =
"logit",set.seed(38902))
matching

# muestra emparejada 56x56
#matching<-matchit(data=datos,group~gender + age + civoc +
smoke,caliper=1.5,
                  #method ="nearest",ratio=1,distance =
"logit",set.seed(38902))
#matching
```

```

# muestra emparejada 40x40
#matching<-matchit(data=datos,group~gender + age + civoc +
smoke,caliper=0.61,
                    #method ="nearest",ratio=1,distance =
"logit",set.seed(38902))
#matching

# muestra emparejada 34x34
#matching<-matchit(data=datos,group~gender + age + civoc +
smoke,caliper=0.2,
                    #method ="nearest",ratio=1,distance =
"logit",set.seed(38902))
#matching

#summary(matching)

## PASO 2: OBTENCIÓN DE LA MUESTRA EMPAREJADA
## ++++++
muestra<-match.data(matching)

## PASO 3: REPRESENTACIÓN GRÁFICO DEL EMPAREJAMIENTO
plot(matching,type="jitter",col="blue")

##=====
##+++++
##=====

```